BRIDGING THE GAP: PROBING THE EFFECTS OF DNA METHYLATION ON BINDING DYNAMICS THROUGH NOVEL METHODOLOGY, AND HISTORICAL REVIEW

By
Gabriel Deards

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Biomedical Engineering

Baltimore, Maryland
December, 2015

Abstract:

Though more than a decade has passed since the completion of The Human Genome Project, the knowledge derived from it has had little impact on our understanding of human health. The reason for this knowledge gap is that only 6.5-10% of the human genome appears to have direct functionality, with the remainder of functional genome elements existing hidden within “junk DNA”. Much effort has been directed towards identifying such regulatory components, like the ENCyclopedia of DNA Elements (ENCODE) project, however, such studies are incapable of distinguishing specific from non-specific protein-DNA binding, or intrinsic vs. environmentally affected binding affinity.

A new method of producing randomly methylated DNA libraries is presented here which should help bridge this knowledge gap. Using mobility shift assay (EMSA), cleavage-qPCR, and immunoprecipitation sequencing (IP-Seq), this method was used to probe the activity and binding strength between Type II restriction endonucleases and randomly methylated DNA. The generation of randomly methylated libraries of genetic material will enable future studies of protein binding affinity, relative to both genetic sequence, and the presence of local and distant DNA methylation. Such experiments can integrate further elements of epigenetic complexity into current methods for identifying binding motifs of proteins, and provide further insight into the underlying mechanisms of epigenetic modification.

Thesis Supervisor: Winston Timp

Title: Assistant Professor, Biomedical Engineering

Acknowledgements:

I want to thank my thesis adviser, Winston Timp, for his guidance and limitless patience in dealing with me. I'd like to thank my friends and family for being supportive and accommodating my bizarre work hours and lengthy periods of radio silence. Lastly, I'd like to thank my peer reviewers and faculty editors, as I feel it's wise to butter them up early.

Table of Contents:

List of Figures and Tables:

FIGURES PAGE

TABLES PAGE

Chapter 1 – History of Epigenetics:

1.1 – The Human Genome Project

The Human Genome Project was a massive undertaking for genetic science. To date, it remains the largest collaborative biological effort in history, requiring almost two decades of work, and collaboration between twenty universities across the globe[1]. This effort, while unprecedented in scope, promised benefits that justified such an investment. The theory went that by identifying functional elements within the genome that guided protein expression, scientists would have developed a tool that could accurately predict an individual's phenotypic response to practically any biological stimulus simply by examining their DNA.

Such a tool would be invaluable to the fields of biological and medical research. By identifying oncogenes, for example, doctors could diagnose likely cancer onset in their patients far in advance of the development of dangerous symptoms. These doctors could then prescribe new personalized drugs, designed specifically to counteract their cancer, potentially halting the spread of the disease entirely. Through further research, these techniques could be applied to other diseases as well, paving the way for a myriad of effective new drugs with boosted potency, and greatly reduced risk of undesired side effects.

Moreover, the benefits of Human Genome Project extend beyond human biology. Using the lessons learned in completing the massive undertaking, mapping out the functional genomic elements of other

organisms would be far less time consuming, and labor intensive. Generating similar genomic mappings for common crops and livestock for example, could pave the way for breakthroughs in agricultural output, the development of cheap biofuels, as well as synthetic biological development, and much more. In effect, unlocking the human genome would crack the mystery of the central dogma of molecular biology, letting scientists read the blueprints for all life, and theoretically, make any alteration they pleased.

Upon its official completion in 2003[2], two years ahead of schedule, the scientific world held its breath in anticipation, confident that the inevitable string of new breakthroughs was soon to follow. However, such developments have been slow in coming, for while scientists did indeed have a complete genetic catalogue of the human genome, scientists have struggled to interpret this new data and identify functional elements within the genome.

It took little time to establish that roughly 99% of nucleotides in the human genome didn't appear to directly code for proteins[2].Later studies determined that the majority of conserved regions (in mammals) consisted of non-coding regions[3,4,5,6], and that most loci associated with phenotypic expression were found within non-coding regions as well[7,8]. These studies and others suggested the inescapable conclusion there must exist functional elements within noncoding regions of the genome,

stored not directly in the genetic code, but in tangential *epigenetic* alterations external to DNA.

The relationship between these epigenetic mechanisms and gene expression has been intensively studied for the past thirty years. Since the discovery that DNA methylation had an impact on gene regulation35 below[35], research has expanded to encompass several other epigenetic modifications, as well as the focused study of complete epigenomes. In recent years, there have been substantial efforts to characterize changes within the human epigenome, including The Cancer Genome Atlas (TCGA) project, The Epigenetics Roadmap[9], and The ENCyclopedia of DNA Elements (ENCODE)[3].

Common human disease's and the methylation state of DNA have been linked through this type of work,[10,11] though the general functional mechanism of methylation remains elusive. Despite present efforts, there remain several key obstacles to overcome in order to characterize the specific effects of methylation. At present there is no clear correlation between the location and or degree of methylation change and gene expression, so determining areas where methylation has a direct causal effect remains elusive.

Further confounding efforts is the lack of high-throughput methods that can determine both the specific location of protein binding, as well as the strength of each particular bond to DNA. The most modern methods of examining protein-DNA complexes currently utilized by

ENCODE, like formaldehyde cross-linked immunoprecipitation, are designed primarily to establish where protein has bound onto genetic material, and can only determine the strength of particular bonds under certain limited circumstances. This deficiency makes it surprisingly difficult to characterize and differentiate relevant binding motifs, resulting in many studies relying heavily on complex biostatistical analysis in order to make sense of data.[12]

On the other end of the spectrum, other classical methods of probing protein-DNA binding dynamics also fall short. Studies have shown that electrophoretic mobility shift assays (EMSA)[13], DNA footprinting[14], and surface plasmon resonance (SPR)[15] also have difficulty determining the strength of protein binding to DNA. Even single-molecule analytical techniques like atomic force microscopy (AFM)[16] which have been used to determine binding strength of protein-DNA complexes in the past, are ill suited to tackling this problem, as they do not scale well to high-throughput analysis and typically require expensive or complicated experimental setups to obtain results.

What all of this suggests is that currently existing tools are insufficient to tackle this problem. High-throughput methods are essential for modern biological research, but if they only generate simple characterizations of epigenetic effects without biophysical insight, they stand little chance of establishing the underlying mechanisms of epigenetic regulation. With a greater understanding of how and where

methylation impacts protein binding to DNA, we can potentially develop experimental tools and methods which would substantially reduce workload by determining which methylation mutations are significant. Once such functional elements have been identified, studies can be undertaken to determine how specific instances of methylation effect gene expression at varying points in the genome, granting us insight into how such a result might be ameliorated, or even reversed.

If present suspicions prove correct, and methylation outside of motifs or outside of cognate binding sites does indeed have a direct effect on their expression, a potentially missing link may be found in the fundamental dogma of molecular biology, and the promises of the human genome project may yet come to fruition. It is with these possibilities in mind that we intend to study methylation in the context around a putative binding site as well as within the binding site of well examined and understood protein-DNA interactions, as well as highlight the specific methods with which further studies could be completed. Through standardizing these methods, we hope to pave the way for a swifter development of further epigenetic study, as well as a deeper understanding of the biophysical ramifications of epigenetics.

1.2 – Epigenetics vs. Genetics

The precise meaning of the term 'epigenetics' was a troublesome question for the scientific community during the 20th century. By necessity, epigenetics evolved rapidly in response to our increasingly sophisticated understanding of genes and gene expression.

The first of these was Conrad Waddington, a researcher knowledgeable in both genetic and developmental biology, who originally coined the term epigenetics in 1939[17]. In early stages, epigenetics was broadly defined as the study of the genetic program for development, as it proceeded step by step. Applying this idea macroscopically, he developed the term epigenotype, which he defined in even broader terms as "The total developmental system consisting of interrelated developmental pathways through which the adult form of the organism is realized."[18] Despite the considerable breadth, and absence of depth of such definitions, epigenetics remained a topic of intermittent discussion amongst biologists[19,20], seeing as not all heritable traits could easily be explained purely through genetics.

In 1996, Russo et al. defined epigenetics as "[T]he study of mitotically and or mitotically heritable changes in gene function that cannot be explained by changes in DNA sequence."[21] What is notable about this definition, as well as the field of epigenetics itself, is how it is cast explicitly as the necessary obverse of genetics. The field studies a vital, and formerly missing piece of the puzzle of heredity, exploring how

various chemical modifications *external* to DNA can alter how information in cells is stored, transmitted, and ultimately expressed as phenotype, and bridge the gap between genetics and developmental biology.

This was not always the case however. Following the discovery of the molecular structure of DNA by Watson and Crick in 1953 establishing it as the primary vector for genetic information, epigenetics was recast as a foil of traditional genetic research, focusing on heritable changes that did not result directly from coded DNA. Little progress was made in identifying such changes however, and as a result epigenetics played second fiddle to genetics research for the remainder of the 20th century.

With the publication of more sophisticated genetic models however, particularly those concerning DNA methylation and its impact on gene expression, epigenetics began to pick up speed, with an enormous amount of progress being made in the past 30 years[18]. It became increasingly apparent that the complexity of gene expression cannot be understood through the lens of genetic study alone. In order to fulfill the promises of the human genome project, a more directed examination of epigenetics, and the epigenome has been undertaken.[3] How this came to be however, can be better understood with an examination of how epigenetics emerged as a field, and its historical role in the development of biology.

1.3 – Unanswered Questions

Though the relationship between genetic code and developmental gene activity was clarified immensely during the mid-20th century, it became increasingly difficult to ignore several mechanisms that defied traditional genetic explanation. A basic but biologically fundamental example would be that of differentiated cells like lymphocytes and fibroblasts maintaining phenotype through cycle after cycle of cell division. It was understood that all cells from the same organism contained the same genome, and while methods of gene regulation were understood as early as 1961[18], these were not sufficient to explain how the expression pattern of these cells could be stably maintained over several generations. Within each cell type, the specialized genes determining its phenotype were somehow permanently turned on (with the remainder of specialized genes permanently turned off), and passed on heritably to further generations without apparent genetic cause.

Stem cells posed similar problems. While such cells are originally unspecialized, when they divide, a new differentiated cell is produced along with another undifferentiated cell. In other words, stem cell division results in an alteration of gene activity of the subsequent specialized cell, despite all having the same genetic input. Additionally, they are capable of long-term self-renewal[22], meaning stem cells can self-replicate and remain unspecialized for years *in vivo*, depending on the type of cell. Considering the difficulty scientists have had in isolating and

growing such cells *in vitro* without having them differentiate, with a gap of nearly twenty years between the isolation of mouse and human embryonic stem cells[23,24], such a mechanism is all the more remarkable.

Yet another example would be the X chromosome in female placental mammals. In 1959, Ohno *et al.* demonstrated that X chromosomes in female mammals appeared to be different, with one highly condensed and heterochromatic chromosome, and another quite similar to the other autosomes in the cell.[25] Females possess two X chromosomes in each of their cells. During development, one of these two chromosomes is permanently and randomly inactivated, leaving the remaining X chromosome active in the cell. Gene expression is altered permanently, and passed on hereditarily, once again external to the currently existing genetic model. Reductively, these examples must therefore be attributable to underlying epigenetic mechanisms; though what these might be remained unclear for quite some time.

1.4 – The Importance of Methylation

Scientists were aware of DNA methylation as early as the identification of DNA as a genetic vector[26,27], though its significance at the time was poorly understood. Many scientists had proposed that methylation might have a role in gene regulation[28], but it would take over thirty years for any evidence to emerge supporting this claim.

DNA methylation is one of the key means of epigenetic modification and expressive regulation found in biology. Methylation acts upon DNA in a variety of different ways, with alternate mechanisms having been recorded in plants, fungi, bacteria, and vertebrates, the characteristics and biological uses of which vary drastically.[62]

DNA methylation is a covalent modification of DNA, which only occurs on a cytosine base in mammals. Additionally, when such cytosines are methylated, the single methyl group ($-CH_3$) is positioned carbon 5, generating 5mCytosine. Most often, DNA methylation occurs at CpG dinucleotides, which themselves can be fully methylated if both strands are modified, or hemi-methylated if only one strand has a methyl modification. These configurations can be better described as symmetrically or asymmetrically modified, as it illustrates more clearly the importance of the palindromic properties of CpG dinucleotides.

The majority of CpG dinucleotides found in the human genome are methylated under normal circumstances, and those left unmethylated tend to be found bunched together in what are known as ‘CpG islands’.

The methylation state of regions of CpG dinucleotides has been found to be greatly significant, with much study devoted to the relative distribution of CpG islands in the genome, and their tendency to be found within the promoter region of genes[29,30]. Observations noting that unmethylated CpG islands are associated with promoters of important genes expressed in most cells, called 'housekeeping genes', led to the hypothesis that methylation played an important role in regulating gene expression.[31]

1.4.1 – Evolution and Function of Methylation

Far from a novel mechanism, Methylation is an ancient phenomenon, having first evolved in prokaryotic organisms as a means of defending against foreign DNA coming from viruses, etc. by relying on specific DNA methyltransferase enzymes, and complementary restriction enzymes.[32,31] The methyltransferases would scan the host organism's DNA, and selectively methylate the DNA at short, palindromic sequences. Restriction enzymes would cleave this same sequence of DNA, only if it was found to be unmethylated. These two groups of enzymes working together allowed for bacteria and other prokaryotes to be able to distinguish host from intruder DNA, and quickly destroy it before it could hijack the organism's cellular machinery.

Methylation works exceedingly well in this capacity for prokaryotic cells, but would present some issues if it were to perform in the same manner for eukaryotes. In eukaryotic organisms, the mechanisms of

methylation have been repurposed to allow dividing cells to inherit states of gene activity, as well as act in several epigenetic regulatory processes active in all eukaryotic organisms.[62]

As stated, cytosine methylation is associated with several epigenetic phenomena: The first of these, which is widely conserved and likely represents the original ancestral role of DNA methylation, is "genome maintenance," where methylation of repetitive elements of the genome enables the control of genes and mobile elements (primarily transposons)[33].

As it happens, mammalian and plant genomes contain more transposons than genes overall, and the majority of DNA methylation in their genetic sequences tends to be found concentrated at these transposons. The relatively high density of methylation serves a very important purpose; as it suppresses the transcription or transposition of transposon sequences, preventing their propagation through the genome. This mechanism is similar to that found in prokaryotic organisms, in that it offers a means of protecting against genomic parasites.[30]

Two papers were published in 1975, proposing a molecular model that could explain the features of complex organisms mentioned above, based on the enzymatic methylation of cytosine nucleotides in DNA. These proposals had a great deal in common with one another (remarkable, considering how they were working independently at the time). Both Riggs et al.[34] and Holliday and Pugh et al.[35] argued that the

methylation of cytosines in mammalian DNA could have a significant effect on gene expression, and that changes in methylation state could be the key to explaining how genes are activated or inactivated during development.

Both observed the propensity for methylation to occur at CpG doublets, which can be thought of as short, genetically palindromic sequences. This is because the reverse transcription of CG is CG, when 5'-3' directionality is observed.

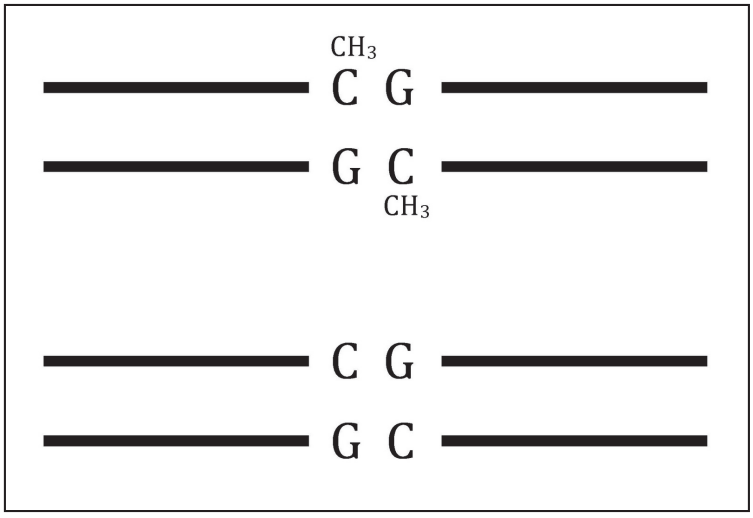


Figure 1: Configuration of a CpG doublet in double stranded DNA

To make such methylation states heritable however, a further mechanism must be involved. During DNA replication, a fully methylated strand of DNA is unzipped, resulting in the separation of the 5' and 3' strands. These methylated, single stranded DNA molecules can then act as templates for the creation of duplicate DNA molecules. Polymerases would add new unmethylated dNTP's to these daughter molecules, resulting in strands that are hemi-methylated (or asymmetrically methylated). If an enzyme existed, they argued, which could completely methylate strands of hemi-methylated DNA, then

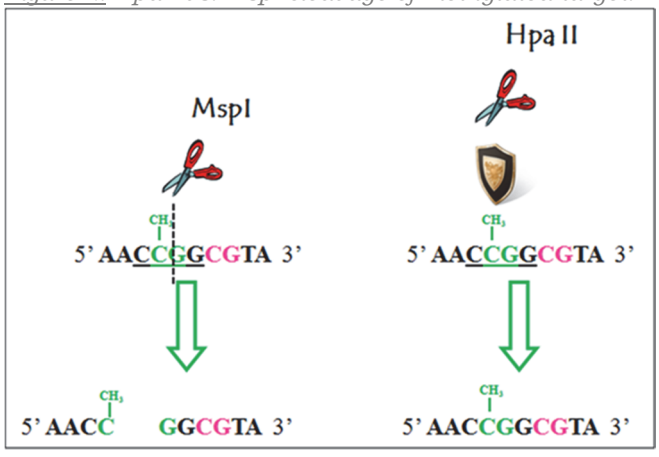
methylation states could be stably inherited, and potentially serve as a vector for gene expression.

1.4.2 – Methylation and Restriction Enzymes

Their arguments were elegant, but entirely theoretical, as neither had proof for the effect of methylation on gene activity. But following that year, an increasing body of evidence began to emerge concerning cancer cell gene expression and mutation, and both Riggs and Holliday argued that methylation might play a part in the malignant tendencies of cancerous cells.[36,37] The development of new molecular biological techniques such as molecular cloning[38] and DNA sequencing[39], gave researchers the tools necessary to screen for methylation state in specific DNA sequences. Neither of these techniques would be possible, however, without restriction enzymes[40].

Restriction enzymes, or restriction endonucleases, are specialized enzymes capable of cleaving DNA at highly specific recognition sites. They do this by making an incision into both strands of the sugar phosphate backbone of the DNA double helix, and their reliability and precision makes them invaluable tools for gene cloning, and protein expression experiments. Of particular importance to epigeneticists were subsets of

Figure 2: HpaII vs. MspI cleavage of methylated target.



restriction enzymes that shared the same target sites, but responded differently depending on the DNA's methylation status. HpaII and MspI are examples of such endonucleases, known as isoschizomers. Both enzymes would target and cut at the sequence 5'-CCGG-3', cleaving between the two cytosines on either end of the DNA strand; however when the second C is methylated, HpaII can no longer cleave the DNA.

Using such isoschizomers, researchers were able to determine the methylation status of promoter regions of the genome, noting in the majority of cases that highly methylated genes were inactive, while unmethylated genes remained active.[41]

1.4.3 – Methylation, and Disease Studies

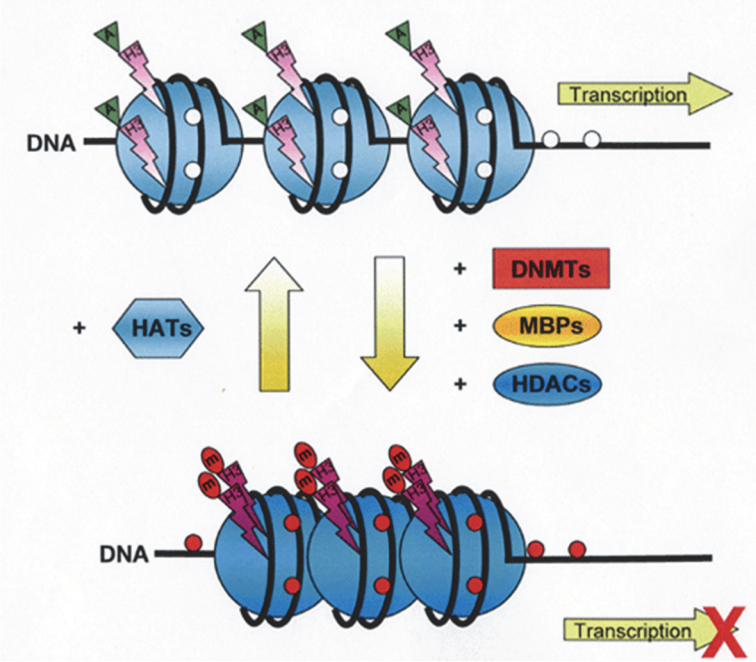
Such discoveries lead to an explosion of activity, as the presence of a more subtle form of inheritance not connected to sequence changes in DNA. Of particular importance was the work of Feinberg and Vogelstein, who in 1983 published a study connecting hypomethylation of genes in cancer cells to the presence and metastasis of cancer in humans.[42] Further studies noted that hypermethylation of identified tumor suppressor genes had similar oncogenic effects.[43] These domains can increase cellular plasticity and shifts in gene expression with disastrous effects for the cell.

These and other discoveries in the 1980's could not be adequately explained by currently existing genetic models, but researchers were slow to think of them collectively as results of epigenetic change. Robin

Holliday is credited for resurrecting / re-popularizing the term epigenetics in the 1990's with his paper "The inheritance of epigenetic defects" published in 1987[44,45]. Soon afterwards, John Maynard Smith coined the term "Dual Inheritance", establishing the existence of classical genetic inheritance, based on changes to DNA sequence, and epigenetic inheritance, classified as anything not based on changes to DNA sequence. As the 21st century dawned, genetics, gene function, and epigenetics exposure have increasingly become integrated in the process of studying cancer, as well as other diseases like diabetes, autoimmune disease, and even aging.[46]

1.4.4 – Beyond Methylation

Additional epigenetic mechanisms beyond DNA methylation have also been explored. For example, chromatin structure and its interface with gene expression has emerged as an active field of epigenetic research[48]. Chromatin is the complex of DNA, RNA and protein found in cells, and depending on its conformational structure, can be accessible, or inaccessible to transcription machinery.



Evidence has emerged suggesting that histone modification, via acylation and methylation, play a crucial role in switching between open and closed chromatin states[47].

Figure 3: "Chromatin Modifications and Their Function." Cell 128, no. 4 (February 23, 2007): 693–705. doi:10.1016/j.cell.2007.02.005.

The mechanism by which chromatin configuration can be inherited is still debated; however the evidence that DNA methylation is the source of shifting chromatin states is compelling[18,47,48]. Further research regarding the potential role of RNA in epigenetic events has also been proceeding rapidly in recent years.

1.5 – The Human Epigenome Project

As stated before, the Human Genome Project was an unprecedented milestone of molecular biology, yet it ultimately provided more questions than answers for biologists. With this lesson in mind, the epigenome project has been launched, aiming to identify the pattern of DNA methylation in numerous cell types. While utilizing numerous analytical techniques to this end, the project ultimately hinges on the usage of bisulfite sequencing, in order to check for methylation at every

relevant site in DNA. While it has been vastly improved since first developed in 1992 by Frommer *et al.*[49], the chemistry underlying the mechanism remains unchanged.

Bisulfite ions actively deaminate cytosine nucleotides, converting them to uracil, a RNA nucleobase. However, the additional methyl group in 5-methylcytosine protects it from nucleophilic attack, leaving them unaffected by the bisulfite ions. When this bisulfite-treated DNA is amplified, DNA polymerases will recognize the 5-methylcytosine as cytosine, and the uracil nucleobases as thymines, and create new strands accordingly.

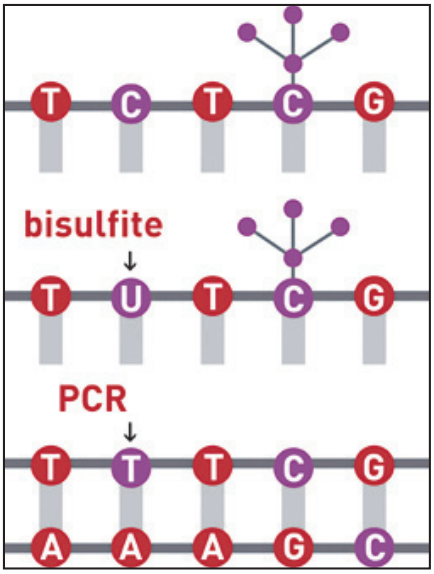


Figure 4 Bisulfite Sequencing Image. Digital image. Bisulfite Conversion. <https://www.diagenode.com/applications/dna-bisulfite-conversion>.

The amplified DNA can then be sequenced, and compared with the original, non-methylation sensitive sequence data to determine which cytosines were methylated (remaining C's in the bisulfite data) or unmethylated (recorded as T's in the bisulfite data). While data is being produced continuously concerning the methylation status of various cell types, the underlying mechanisms behind the effects of DNA methylation on gene regulation still remain poorly understood. In order to derive the most benefit from the ongoing epigenome project, much work remains towards this end.

Chapter 2 – Characterizing Epigenetic Effects:

2.1 – Methylation and DNA

Despite having considerable impact on gene expression, the effects of methylation on DNA structure is surprisingly subtle. Some aspects of why this is so are counterintuitive, but can be better understood by examining the tertiary structure of a DNA strand.

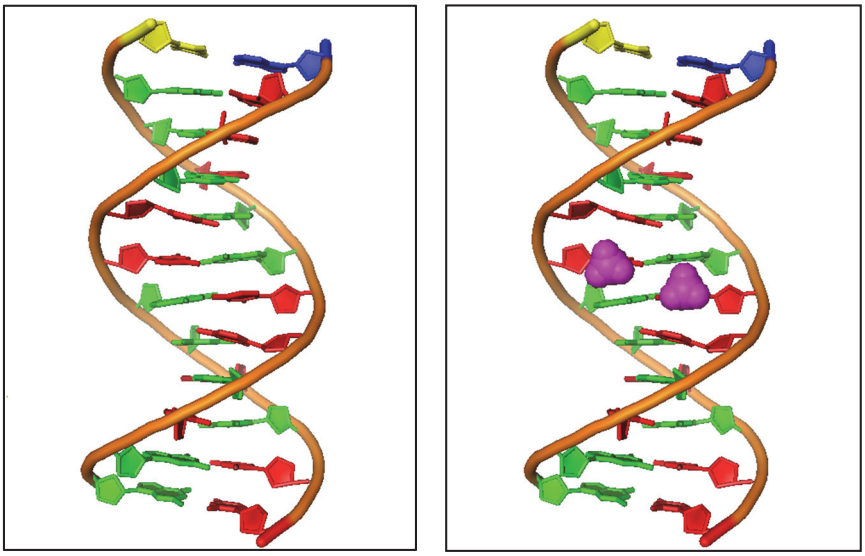


Figure 5: Unmethylated and Methylated Strands of DNA.

These PyMOL images represent the same strand of DNA (sequence 5'-ACCGCCGGCGCC-3')[50], chosen due to its having an internal CpG doublet, which is completely methylated in the second image. The DNA itself has been color coded according to nucleotide (A-> azure, C-> crimson, G-> green, T-> yellow), with methyl groups shown as space-filled models in magenta.

Comparing the methylated and unmethylated sequences, several facts become apparent. For one, the presence or absence of methylation

does not appear to modify base pairing, as both C and 5mC (methylated cytosine) pair with G in DNA sequences without interference from DNA mismatch repair[51]. Additionally, methylation has a very minor effect on the shape of the DNA double helix, with the additional methyl groups residing comfortably in the major grove of DNA, resulting in only minor conformational change.

Considering the demonstrated effects that epigenetic modification can have on gene expression, this result is somewhat underwhelming; and in general, the molecular mechanisms explaining the function of DNA methylation are poorly understood[52]. However, several discoveries have recently emerged that can potentially explain the functional impact of cytosine methylation, though further experimental verification remains necessary. Kinetic experiments and Monte Carlo simulations have shown that while methylation does not affect DNA bend magnitude and direction, it can result in a loss of DNA flexibility, as well as subtle under-winding of methylated CpG dinucleotides[53,54].

2.1.1 – Methylation Distribution

The distribution of CpG dinucleotides in the mammalian DNA is also a source of interest for researchers. The human genome is roughly three billion base pairs in length. Assuming an equal frequency of DNA bases (P(A) = P(C) = P(G) = P(T) = ¼), we should expect that CpG dinucleotides should comprise roughly 1/16th of the genome, or 187,500,000 dinucleotides. In reality however, CpGs are actually under-

represented, with only about forty million in the human genome. Of this smaller number, 80% are methylated at any given time,[55] making the total number of methylated CpGs around forty million.

In healthy human cells, these CpGs are distributed unevenly throughout the genome, concentrated primarily in areas known as CpG islands. These islands, as previously mentioned, commonly correspond to the location of promoter regions in DNA. Methylation of these sites is also uneven, found primarily within gene bodies, inter-genic regions, and in repeated sequences, while those found at CpG islands are typically unmethylated. This is not universally true, as many some promoter regions like OCT4 have been found to be methylated in somatic cells[56,57,58]; but for the most part CpG islands are free of methylation.

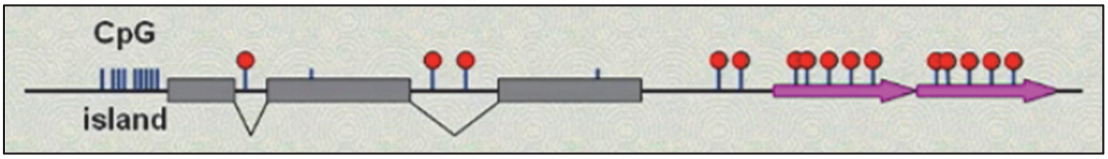


Figure 6: Methylation Pattern of a healthy cell.
CpG doublets are shown as blue lines, while additional methyl groups are shown as red dots.
Defossez, P. (2011), "Proteins that Bind Methylated DNA", DNA Methylation: Physiology, pathology and disease*, The Biomedical & Life Sciences Collection, Henry Stewart Talks Ltd, London*

The methylation patterns in cancer cells are markedly different from healthy cells. The total amount of methylation goes down sharply, primarily in the repeat sequences of the genome.[59,60] Conversely, some formerly unmethylated CpG islands become hypermethylated.

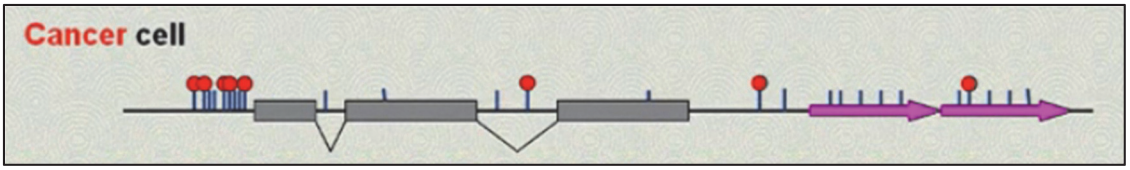


Figure 7: Methylation Pattern of a Cancerous Cell.
In cancer cells; CpG islands become hypermethylated, while total amount of methylation decreases.
Defossez, P. (2011), "Proteins that Bind Methylated DNA", DNA Methylation: Physiology, pathology and disease*, The Biomedical & Life Sciences Collection, Henry Stewart Talks Ltd, London*

The genetic basis of cancer is driven by the deactivation of oncogenes, and loss of tumor suppressor genes like p53, etc.[61]. These defensive genes come in pairs, with two copies of each per cell. If both genes are lost for purely genetic reasons, due to point mutations, or loss of heterozygosity, transformation can occur and cancer may develop.[62] However, these genes can also become lost for epigenetic reasons, as is the case when promoters become methylated, for example. DNA methylation therefore is tightly linked to cancer.[61,55]

2.1.2 – Protein Interactions

Proteins, such as many transcription factors, that bind to specific DNA sequences are essential for the proper regulation of gene expression.[63] Identifying the specific sequences that each factor binds can help to elucidate regulatory networks within cells and how genetic variation can cause disruption of normal gene expression, which is often associated with disease. Traditional methods for determining the specificity of DNA-binding proteins are slow and laborious, but several new high-throughput methods can provide comprehensive binding information much more rapidly. Combined with *in vivo* determinations of

transcription factor binding locations, this information provides more detailed views of the regulatory circuitry of cells and the effects of variation on gene expression.

Numerous studies have shown the regulatory effect DNA methylation can have on gene expression.[68,68] Despite this recognition however, the mechanisms that drive these effects are themselves poorly understood. The reason for this is due in part to the contextual nature of these effects, and the difficulty scientists face in establishing what effects are in play.

The presence of a methyl group itself represents a relatively small chemical alteration to DNA, but when added to cytosine, the biophysical properties of the nucleotide are changed sufficiently to affect its base readout.[64] The presence of additional hydrophobic contacts alters cytosines presence in the major groove of DNA, while simultaneously altering its electrostatic profile and physical shape with regards to the minor groove.

Recent evidence suggests that CpG methylation can increase DNase I activity and cleavage potential at bordering positions by orders of magnitude, simply through the resultant physical alterations to DNA 3D structure.[65] This discovery, when considered in the context of the regulatory properties of CpG methylation, suggests a potential explanation of methylation sensitivity for transcription factors, and protein-DNA binding affinity. However, this potential explanation cannot

be considered in a vacuum. The biophysical qualities that result in the behavior of CpG methylated DNA also result in alterations to nucleosomic stability, which has enormous downstream effect on a long list of relevant epigenetic factors, such as local and global chromatin structure.[66]

The disappointing conclusion to draw is that when the mechanisms underlying DNA methylation's regulatory properties on genetic expression are revealed, there will be few absolute rules that will guide the process. Due to the complexity of the underlying forces which drive the process, the epigenetic context of CpG methylation, rather than the content itself, will be what determines its behavior.

2.2 – Methyl Binding Proteins

While the link between DNA methylation and gene regulation is well categorized, and extensively studied, the underlying mechanism behind its operation is poorly understood. Numerous studies have established that DNA methylation can result in transcriptional silencing in mammalian cells. This is made possible by two separate methylation sensitive effects working in tandem. Firstly, DNA methylation can prevent certain transcription factors from binding to their specific target sites. An example of this can be seen with chromatin insulator protein CTCF, which is blocked from binding by DNA methylation[67,68]. At the same time, DNA methylation creates new binding sites for proteins which specifically target DNA, known as Methylated DNA binding Proteins

(MBPs). These proteins are an essential aspect for creating and maintaining a transcriptionally silenced chromatin environment[69].

Methylation binding proteins were actually discovered by accident by Adrian Bird et al. in 1989[70]. Ironically, their initial intention was to identify factors which bind to unmethylated DNA, thereby protecting adjacent CpG islands from methylation,[71]. Instead, they discovered the first protein factors, MeCP1 and MeCP2, which bind specifically to methylated DNA sequences, and by isolating the Methyl-CpG-binding domain (MBD)[72], other similar proteins containing the motif were identified via extensive homology search[73,74]. Considering the relative lack of protein databases at the time, the identification of these MBDs was all the more impressive.

Homology can only take you so far however, as similar molecular structure does not necessarily imply similar molecular function. In order to determine whether not a given protein binds to methylated DNA, further tests must be carried out. One of these tests, utilized in the characterization of each MBD is the Electrophoretic Mobility Shift Assay (EMSA) or Gel Retardation Assay.

The gel electrophoresis mobility shift assay is a common affinity electrophoresis technique used to detect and study protein complexes made with nucleic acids, and will be discussed in more technical detail in the experimental section. The scientific principles behind the method are utilized in a wide range of analyses to examine protein DNA interactions,

and has changed little since it was first developed by Fried and Crothers.[75,13]

Compared with similar assays, EMSA is procedurally simple, inexpensive, and capable of supporting a wide variety of different experimental conditions that alter binding conditions. Additionally, when using nucleic acids labeled using either radioisotopes, or biotinylation, the assay is surprisingly sensitive, capable of handling small concentrations and volumes of protein and nucleic acid.[13]

As noted in previous sections, the rate at which an organic molecule propagates through a gel matrix is dependent on how fast it can migrate through pores in the gel. This navigation speed depends largely on the size and charge of the molecule, though shape and flexibility also come into play. All else being equal, according to the principles of electrophoretic separation a protein-DNA complex will migrate at a slower rate than DNA left unbound[76,77]. So, by incubating purified protein with two different ligands (either methylated or unmethylated forms of otherwise identical DNA), and then running mixture through the gel, the protein-DNA interaction can be detected as slowly migrating complexes.

That said, the ability to resolve protein:DNA complexes depends largely on how stable the complex remains during each step of a given procedure. One of the benefits of subjecting protein:DNA complexes to a gel retardation matrix is the accompanying “caging effect”, which helps

slow the rate of dissociation, and enables prompt reassociation of protein with sufficiently high local concentration[78,79]. However, dissociation of proteins can still be relevant experimentally, a topic which will be touched on later.

While EMSA can be used to verify that a protein binds methylated DNA *in vitro*, this doesn't necessarily imply that it will behave the same way *in vivo*. In order to establish whether a particular protein is a member of the MBP family, it must first be determined whether it would bind heavily methylated regions of DNA in living organisms.

Technically there are further questions concerning whether or not proteins bind to heterochromatin, independent of DNA methylation, but such issues can be resolved by running similar experiments in cells with low levels of methylation. Additionally, if a target sequence for your protein has been identified, demethylating this sequence and examining whether or not proteins remain bound can provide further evidence for MBP characterization.

2.3 – Methyl Binding Domains

These approaches and others have been utilized by numerous labs over the past few decades, with an enormous contribution coming from Adrian Bird.[80,81] Despite this work however, researchers have only discovered 9 human proteins that bind specifically and exclusively to methylated DNA. These proteins can be subdivided into separate families, consisting of an original protein discovered biochemically, and

subsequent proteins determined homologically. In chronological order of discovery, these families are: the MBD family, containing MeCP2, MBD1, MBD2 and MBD4, the zinc finger family, containing Kaiso, ZBTB4 and ZBTB38, and the SRA family, containing UHRF1, and UHRF1.

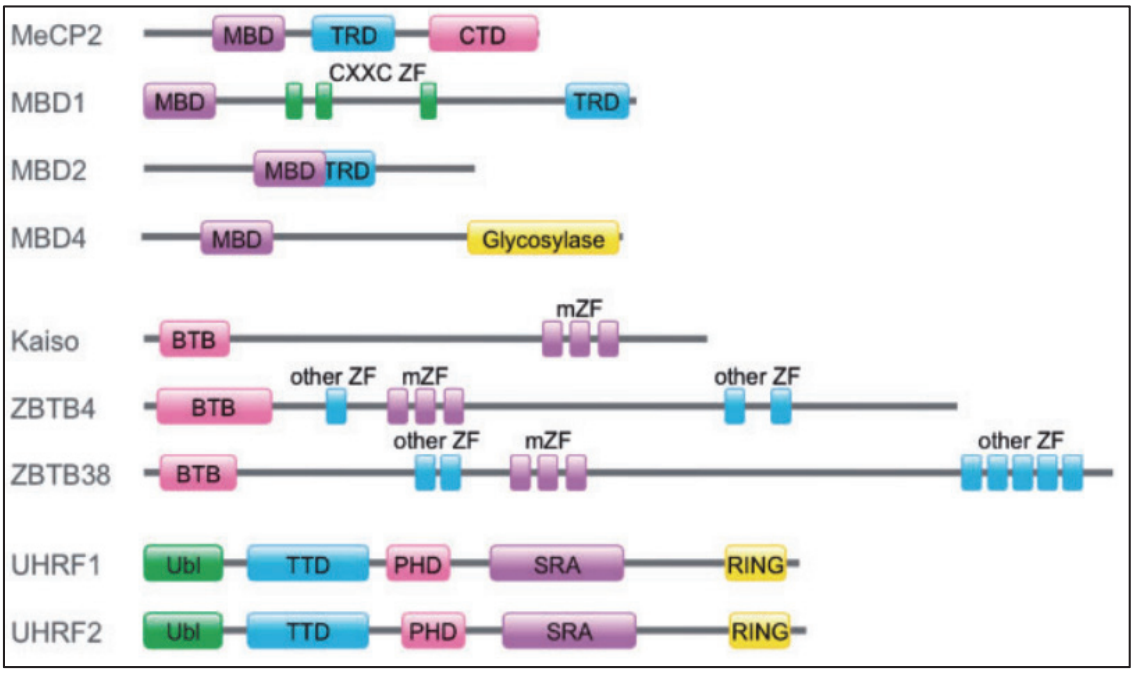


Figure 8: The Three Families of MBD Binding Proteins in Mammals, with Key Domains. Fournier, Alexandra, Nobuhiro Sasai, Mitsuyoshi Nakao, and Pierre-Antoine Defossez. "The Role of Methyl-Binding Proteins in Chromatin Organization and Epigenome Maintenance." Briefings in Functional Genomics 11, no. 3 (May 2012): 251–64. doi:10.1093/bfgp/elr040.

Each family differs significantly in structure, meaning they possess completely different mechanisms for methylated DNA recognition. This implies naturally, that there are several ways to bind methylated DNA, and suggests the possibility that other DNA binding proteins may have similar properties, and remain undiscovered. Determining how these mechanisms work *in vivo* is far more difficult, however. For example, while the solution of the MBD domain bound to methylated DNA has since been determined[82], nothing in the model could explain how MBD

proteins are capable of binding to specific sections of DNA, rather than all methylated DNA, irrespective of the sequence.

The roles of methyl-binding proteins can be examined for further insight into epigenetic regulation. In numerous studies examining each MBP, it has been observed that they initiate heterochromatinization; i.e. they convert the formerly permissive euchromatin to a more repressed, closed off heterochromatin.[83] This alteration appears to be the cause of repressed transmission seen in hypermethylated areas of DNA, rather than being directly caused by DNA methylation itself. Instead, DNA methylation attracts the binding of MBPs, which themselves catalyze the formation of heterochromatin.

2.3.1 – MeCP2

MeCP2, the first MBP discovered, has been strongly associated with the development of Rett Syndrome in humans.[84,85] Rett Syndrome is a severe neurodevelopmental disorder that primarily effects young girls. After mentally progressing normally for their first 18 months, patients with Rett Syndrome rapidly regress, and exhibit severe mental deficiency.

The vast majority of Rett patients are found to have mutations in MeCP2,[86] and experiments on mice have resulted in symptoms that mirror Rett Syndrome, so it is safe to say that the absence of functional MeCP2 results in this sort of neurodevelopmental disorder. However, it remains somewhat unclear why this would be the case, what specific

MeCP2 targets might be involved, or what role methylation has to play in the development of the disorder.

2.3.2 – MBD4

MBD4, also from the MBD family, also has a strong affect on phenotype. When deleted in mice, they appear to be fairly healthy initially; however, they exhibit a greatly increased rate of mutation.[87] When examined in cancer models, the absence of MBD4 increased the development and frequency of tumor metastasis in affected mice. Further study has revealed that MBD4 works actively to repair damage from spontaneous deamination of methylated DNA.

As it turns out, MBD4 is present in methylated regions to combat this issue under normal circumstances. A region of MBD4 has specific thymine glycosylase activity, which can remove offending T's before they are altered by other repair machinery most of the time. While not perfect, MBD4 clearly has a role in limiting the number of point mutations generated in methylated regions of DNA.

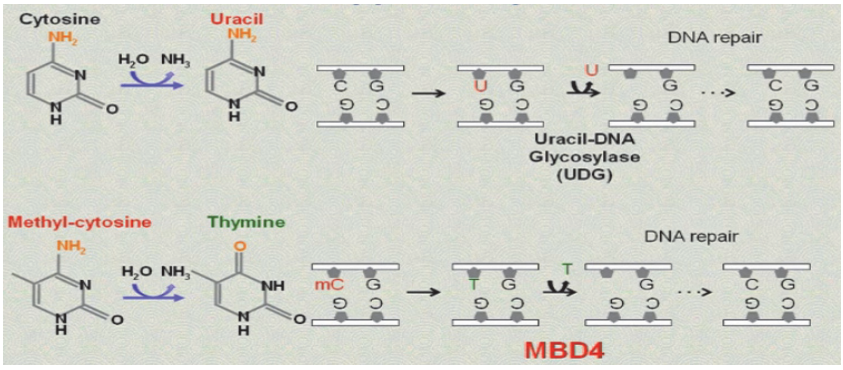


Figure 9: UDG and MBD4 performing separate DNA mismatch repairs.
Defossez, P. (2011), "Proteins that Bind Methylated DNA", DNA Methylation: Physiology, pathology and disease, The Biomedical & Life Sciences Collection, Henry Stewart Talks Ltd, London

2.3.3 – Kaiso

Kaiso is a zinc finger protein, named for the three zinc fingers it utilizes to target methylated DNA, as well as certain unmethylated consensus sequences.[88] Further experimentation will be required to determine whether zinc finger proteins are more strongly drawn to methylated DNA, or sequence specific DNA, as presently researchers are unsure. Much like MBD4, when silenced in mice, there appears to be little sign of phenotype, but when applied to cancer-susceptible mice, such as those of the ApcMin model, the number and size of tumors developed actually decreases.[89]

The reason for this change is currently unknown, but a potential theory involves interaction with methylated cancer suppressor genes. Cancer suppressor genes, which are normally unmethylated and expressive, are found to be abnormally methylated in early stages of cancer metastasis. Such methylation could potentially recruit MBPs like Kaiso, which would bind to the genes, and result in its suppression (and by extension, cancer expression). Operating in an overly methylated environment, Kaiso could potentially operate in a manner detrimental to the host organism, however further substantiation of this theory will be required either way.

2.3.4 – UHRF1

UHRF1, a member of the SRA family, is far more complex than MBD MBPs as can be seen in the image below. Studies in mice have shown that the deactivation of UHRF1 results in death in early embryonic stages of development (and is the only methyl-binding protein with a lethal phenotype).[90] Examination of the deceased embryonic cells showed them to have significantly lower levels of DNA methylation.

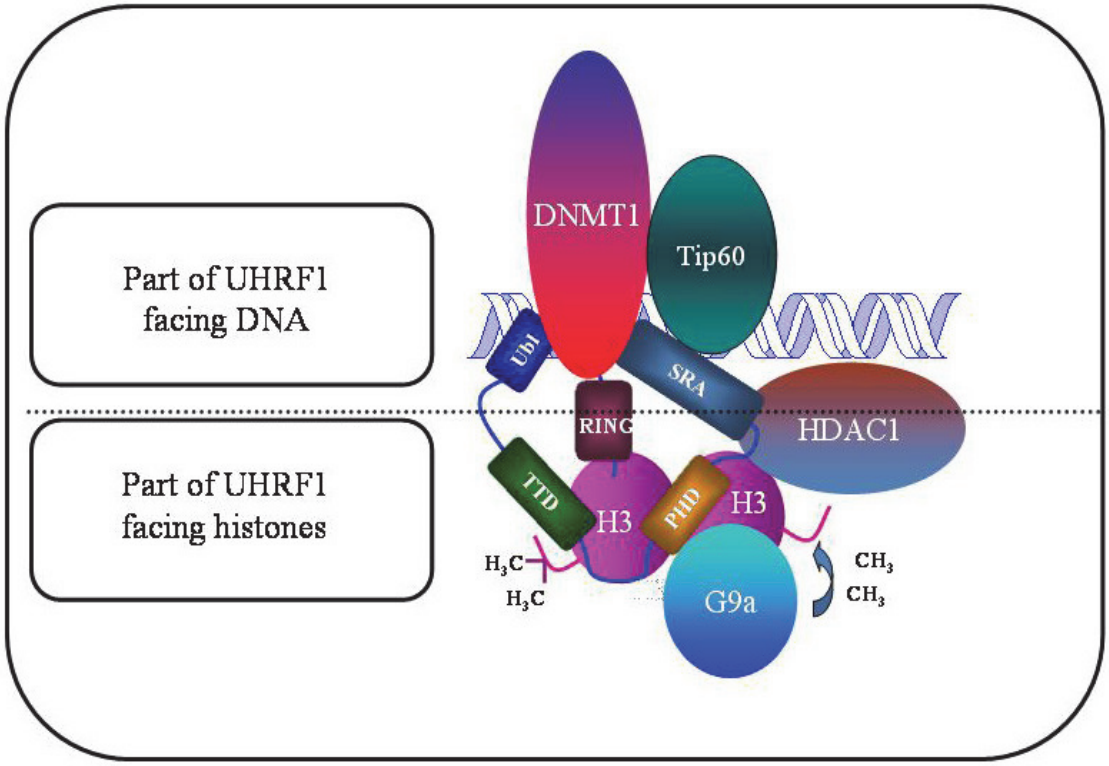


Figure 10: Schematic Representation of UHRF1
Alhosin, Mahmoud, Tanveer Sharif, Marc Mousli, Nelly Etienne-Selloum, Guy Fuhrmann, Valérie B Schini-Kerth, and Christian Bronner. "Down-Regulation of UHRF1, Associated with Re-Expression of Tumor Suppressor Genes, Is a Common Feature of Natural Compounds Exhibiting Anti-Cancer Properties." Journal of Experimental & Clinical Cancer Research 30, no. 1 (2011): 41. doi:10.1186/1756-9966-30-41.

As mentioned previously, symmetrically methylated DNA replication results in two separate strands of hemi-methylated DNA. If

these strands are left to their own devices, and continue to replicate over time as hemi-methylated strands, then the methylation site itself will eventually be lost through dilution as more and more replication cycles take place. UHRF1 has a specific affinity for hemi-methylated DNA molecules, and appears to be instrumental for the maintenance of methylation status after DNA replication.

However, as seen in the examples above, UHRF1 is an exception in this regard. The majority of MBPs currently known appear to be non-essential to their host organism; that is, their absence in living organisms does not result in a lethal phenotype. This is altogether surprising, considering that DNA methylation itself is absolutely vital for eukaryotes. Numerous studies have shown that tampering with DNA methylation enzymes results in lethality in the early embryonic stage of development; so how can this discrepancy be explained?

2.4 – Restriction Modification Systems

As mentioned previously, proteins that cleave at or near specific restriction sites on double stranded DNA. These enzymes, also known as restriction endonucleases, fall into several types and subtypes depending on their specific structure, mechanism, and whether or not they cleave at or near their unique restriction sites. In general, after coming into contact with such a restriction site, restriction enzymes bind, and then sever both strands of the sugar-phosphate backbone of the DNA double helix. This process, known as 'restriction' is a primitive defense

mechanism utilized by bacteria and archaea to defend their DNA from contamination by foreign invasive DNA, destroying it and its ability to alter the host cell.

However, such enzymes cannot operate in a vacuum. Considering the short length of each restriction site, they will inevitably be found within prokaryotic genomes, and if left accessible, these defensive restriction enzymes to would eviscerate host DNA, doing tremendous harm to the host organism. As such, these enzymes have evolved to work in conjunction with a second class of modification enzymes, methyltransferases, which modify prokaryotic host DNA by adding methyl groups to specific base pairs on each strand. The additional methylation alters the restriction enzyme's recognition site, making it impossible to cleave the strands of modified host DNA, thus protecting it from harm.

These processes working together constitute what is known as a restriction modification system, and are essential for the protection of prokaryotic organisms against foreign genetic material. There are several different varieties of restriction modification systems: type I, type II, type IIS, type III, and type IV, named in order of discovery, and categorized according to characteristics of the enzymes involved. Despite their differences however, each system operates on the same principle of restriction enzyme activity being curbed by the presence of methylation,

making such systems, particularly type II restriction endonucleases, of special interest to biologists.

Type II restriction modification systems are both the simplest, and most common genomic defense mechanisms found in prokaryotic organisms.[91] While other modification systems utilize a single molecular complex of bound endonuclease and methyltransferase, Type II modification systems rely on independently operating endonclease and methyltransferase proteins operating in tandem.

The methyltransferase enzyme scans DNA, reading one strand at a time, and adding a methyl group to nucleotides once it has recognized its specific target sequence. These sequences are usually no more than 6 base pairs in length, and are often palindromic. The now methylated strand is then examined by a corresponding endonuclease, which scans both strands of DNA, and responds to the same target site. In the event it comes across this sequence in a methylated state, it will not interact with it in any way. However, when the same sequence is found unmethylated, the endonuclease will promptly cleave the internal phosphodiester bonds holding the strand together, severing the DNA into two broken pieces either at, or very near the target location.

The potential for self-harm on the part of the organism employing this form of modification system still exists, at least theoretically, but is balanced out by thermodynamics. Methyltransferases act as monomers, singular units that are capable of binding to DNA independently, and

acting on singular strands more or less as soon as they come into contact. Type II endonucleases, however, operate as homodimers.

2.5 – Use in Experimental Design

Because the enzymes only act at specific locations, and can differentiate between alternate sequences and alternate methylation states, type two restriction endonucleases provided scientists with an extremely useful tool to utilize in genetic analysis. Sequences could be examined for point mutations, or single nucleotide polymorphisms (SNPs) with near perfect accuracy, which result in either uniform fragments, or an unbroken larger band when subjected to digestion, and visualized via gel electrophoresis. This treatment could be used to quickly establish the nature of a DNA segment without having to use more complex or expensive forms of gene sequencing.

Restriction enzymes, in particular HpaII and MspI, possess several unique qualities that made them ideal candidates for study. Endonucleases bind very stably and specifically to DNA, and remain bound and unable to cleave at their specific binding sites when incubated in the absence of Mg^{2+}. Additionally, HpaII and MspI are isoschizomers, meaning they both target and cleave the same recognition sequence on a strand of DNA (CCGG). However, this is not always the case. When this shared recognition site presents with an internally methylated cytosine, HpaII loses ability to cleave this DNA, while MspI remains unaffected.

Chapter 3 - Experimental Outline:

3.1 – DNA Sequence Isolation and Purification

As a preliminary to any experiments, suitable genetic material had to be created, purified, and tested to ensure an appropriate level of fidelity. To this end, several biotinylated primer pairs were designed, targeting short segments of the promoter of GSTP1 protein in human male DNA. Several sets of primers were designed to target various subsections of the promoter region flanking the GSTP1 gene. GSTP1 has been the subject of extensive study, and specific patterns of methylation in the region have been shown to be strongly correlated with the presence of prostate cancer[92,93]. These qualities made the GSTP1 promoter region an excellent choice for the purpose of testing this new methodology, though theoretically any similar CpG-rich region would do. Planned future experiments will examine the binding regions of CTCF, and MeCP2 for similar reasons.

Under normal circumstances, these primers are unable to interact with DNA, as it is bound in a double stranded configuration. To denature the DNA strands, a thermocycler is used, which raises the temperature of the samples above the T_m or "melting point" of DNA, at which half of DNA molecules are in a random coil configuration, meaning they are single stranded. The T_m for a particular genetic sequence varies dependent on its nucleotidic content, as well as its surrounding solution.

Several suitable primer pairs were generated in Geneious[94] targeting small subsections of the GSTP1 promoter region. Each was designed to generate an amplicon between 100 and 300 bp (base pairs) in length, and to contain at least one binding site for either HpaII/MspI, or BamHI, a methylation insensitive endonuclease used as a control. The primer sets ultimately chosen for experimental verification of this method were thusly named H/M_1 and B_1, and will be the focus of the experimental analysis. Later iterations of these experiments utilized a nested PCR protocol, involving gel images of alternate primer sets can be found in appendix B.

These primers were used in a series of PCR runs, generating amplicons from human male DNA template via separate reactions detailed below. To increase yield of these reactions, as well as to generalize the method for other amplicons, several optimization trials were undertaken in order to determine optimal PCR parameters. The products of these reactions were later purified using Ampure beads according to standard protocol, quantified with nanodrop UV spectroscopy, and checked for transcriptional fidelity via gel electrophoresis and Sanger sequencing.

Several alternate polymerases, thermal recipes, and reaction buffers were tested several in order to find the optimal mixture for amplifying sequences less than 6 kb (kilobases) in length, and possessing a high GC percentage. These initial amplicons were generated via PCR

utilizing human male DNA, unmethylated dNTP's, and OneTaq polymerase as seen in the material and thermal recipes below:

Table 1: Recipe for PCR Experiments.

Reagents	**Volume**	**Final Concentration**
GC Enhanced Buffer	10 µl	1X
10 mM dNTPs	1 µl	200 µM
10 mM Forward Primer	1 µl	0.2 µM
10 µM Reverse Primer	1 µl	0.2 µM
OneTAQ DNA Polymerase	0.25 µl	1.25 units/ 50 µl PCR
Template DNA	1 µl	< 1 µg / 50 µl PCR
Nuclease-free H_2O	To 50 µl	N/A

Table 2: Thermal Recipe for PCR Experiments

Step	**Temperature**	**Time**
Initial Denaturation	94 °C	5 Minutes
30 Cycles	94 °C 45-68 °C 68 °C	30 Seconds 45 Seconds 60 Seconds per kb
Final Extension	68 °C	300 Seconds
Hold	4 °C	Indefinite

The amplicons generated were purified using AMPure beads[95]. This was accomplished by adding paramagnetic beads, suspended in an optimized buffer, to the DNA samples at a ratio of 9:5. The buffer solution selectively binds larger pieces of DNA (greater than 100bp in length), to the suspended beads, while leaving the remaining salts and excess nucleotides in solution.

The mixed samples are then kept on a magnetic rack for several minutes, separating the beads from solution. The beads are then washed twice in 70% ethanol, removing any contaminants that might still remain, and left to air dry for 15 minutes. Lastly, the dry beads are washed in TE elution buffer, separating the PCR product from the AMPure beads, and permitting extraction.

3.2 – Gel Electrophoresis

To establish that PCR produced amplicons of an appropriate size, the purified products were then run through a process known as gel electrophoresis. Gel electrophoresis is one of the simpler methods for separating an analyzing fragments genetic material, based on their charge, and relative size. The process leverages several qualities of nucleic acid molecules in order to achieve this separation. DNA and RNA molecules both contain negatively charged phosphate backbones, giving them an overall negative electrical charge. When these molecules are subjected to an electric field, they are induced to move away from the cathode, and towards the anode.

This quality on its own however, is not enough to effectively separate or analyze nucleic acids. Both DNA and RNA are polymers, consisting of nucleotides that are held together by connecting sugar phosphate backbone. As a result, the overall charge density of such molecules remains constant, as the ratio of phosphates to nucleotides is 1:1. To separate genetic fragments on the basis of size, another of their physical qualities will have to be exploited. Placing the material into a porous polymetric support such as an agarose gel, can cause longer strings of DNA to migrate at an appreciably slower rate despite being subjected to the same electrical field. What's more, the specific relationship between distance traveled, and molecular weight of DNA, can be modeled with great accuracy.

In these experiments, purified DNA and NEB 100bp ladder was combined with NEB gel loading dye. 2% 7x10 cm agarose gels containing LifeTech SYBR Safe DNA gel stain, were made in lab, and then and submerged in 0.1X TAE buffer in a Bio-Rad electrophoretic device. The mixed samples were loaded into the submerged gel, and run at 100 volts over the course of an hour. The gel was then taken out of solution, and imaged under ultra violet light, as seen in the figures below.

H/M_1 and B_1 amplicons generated from human male DNA with using this protocol presented bands corresponding to their molecular weight as compared with the NEB 100bp ladder (see figure below). This ladder contains a heterogeneous cocktail of DNA fragments, designed to

present 12 distinct bands on an electrophoretic gel. The fragments themselves are of known weight and concentration, providing a reference for the estimation of a given DNA products molecular weight.

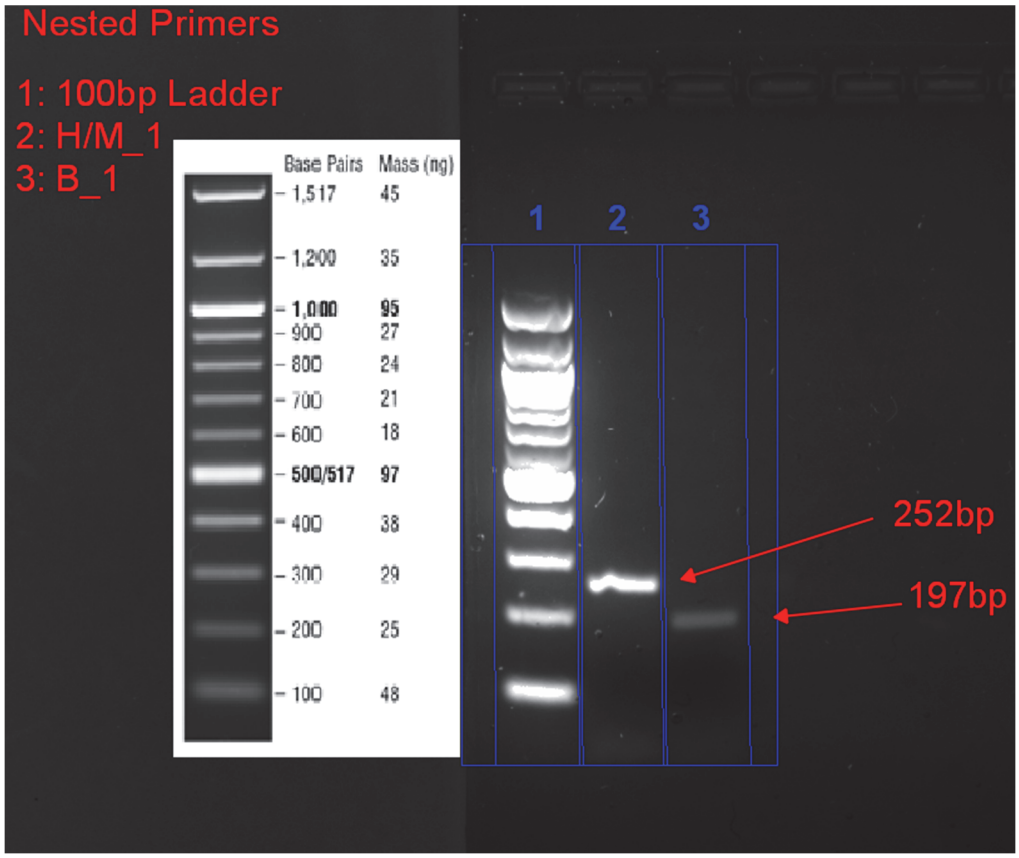


Figure 11: Gel Image of nested DNA samples H/M_1 and B_1.
The size of each sample can be verified by comparing the propagation distance with the DNA ladder in lane 1. The relative weight of each band can be seen to the left of the gel.

3.3 – Sanger sequencing

Sanger sequencing is one of several methods employed by molecular biologists to determine the genetic sequence of DNA samples *in vitro*. Originally developed by Fredrick Sanger in 1977[96], it was one of the most effective and popular methods of DNA sequencing for the remainder of the century, only recently having been overtaken by various Next-Gen

sequencing techniques. Though Next-Gen sequencing has become the standard method for dealing with genome sequencing, transcriptome profiling (RNA-Seq), DNA-Protein interaction (ChIP-Seq), as well as for the generation of epigenomic data[97], Sanger sequencing is still widely used in laboratories to validate generated data.

Sanger sequencing works on principles very similar to traditional PCR. Both require DNA template which is denatured into single strand DNA, appropriately designed primers, a suitable polymerase enzyme, and deoxynucleotidetriphosphates (dNTP's). Where Sanger sequencing differs however, is in the inclusion of chemically modified dNTPs called di-deoxynucleotidetriphosphates, or didNTPs. Like dNTP's they are capable of linking with other dNTP's by forming an ester linkage at their 5' carbon, thus becoming integrated in the new DNA strand. However, they lack a hydroxyl group at their 3' carbon position, preventing other nucleotides from forming a similar bond, thus halting DNA elongation, and terminating the sequence.

Sanger sequencing utilizes this property of didNTPs in the following fashion. The DNA sample to be sequenced is separated into four separate reactions, each one containing all four standard dNTP's, polymerase, primers, as well as one of the four didNTPs, each reaction having a separate didNTP. Crucially, the didNTP is added in roughly 100-fold excess to its nucleotidic counterpart, meaning a concentration of 0.1mM didGTP would be required for a concentration of 0.001mM dGTP.[98]

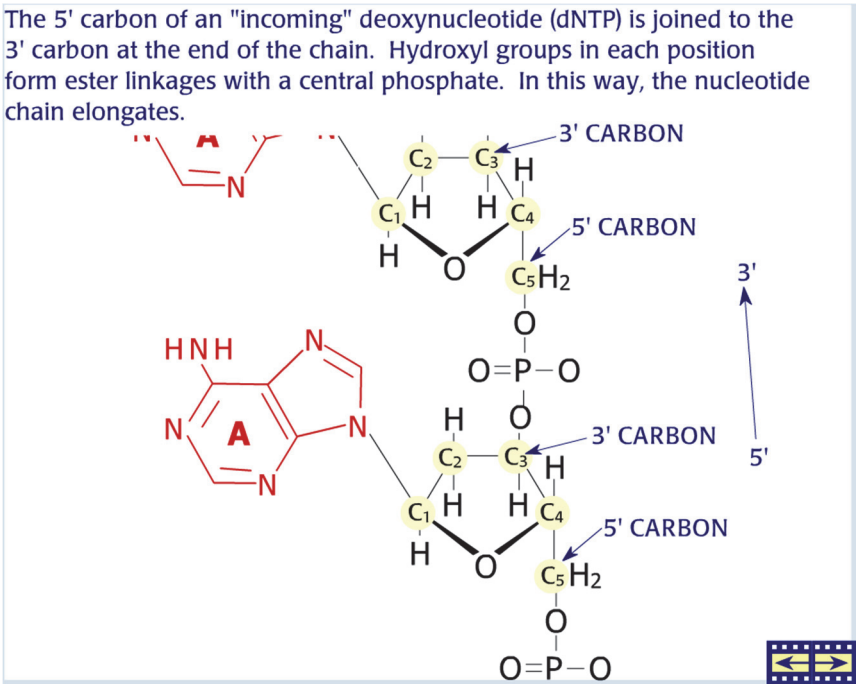


Figure 12: Elongation of DNA Strand Using didNTPs
Nucleotide elongation. Digital image. Early DNA Sequencing. Cold Spring Harbor Laboratory, n.d. Web. 4 Apr. 2015.

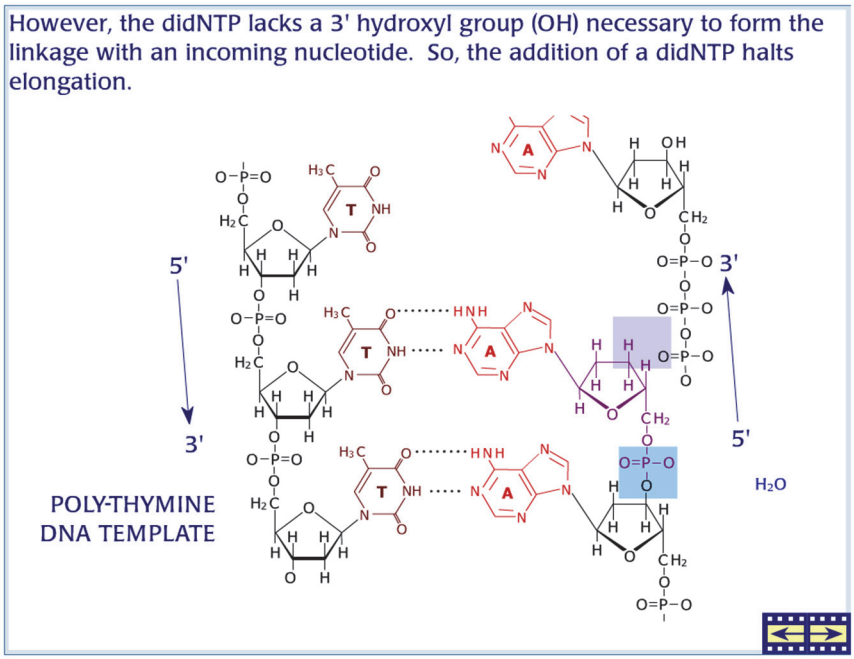


Figure 13: Integration of didNTPs Halts Further Elongation.
Nucleotide elongation. Digital image. Early DNA Sequencing. Cold Spring Harbor Laboratory, n.d. Web. 4 Apr. 2015.

Each of the four samples then undergoes PCR, which with the addition of didNTPs, results in fragments of transcribed DNA of varying levels of completeness, each terminating at a point along the genetic sequence corresponding to that reaction's didNTP. The fragments from each reaction can then be loaded into separate lanes of a denaturing polyacrylamide-urea gel, and subjected to electrophoresis. This results in a series of bands propagating down the gel, each of which corresponds to a fragment of DNA which terminated at a particular point along the sequence. By cross referencing the distance each band has propagated to a molecular weight ladder, the specific didNTP inserted at a specific point along the sequence can be established, and by extension, the entire sequence can be determined.

Since its initial development several technical variations have been employed to increase the experiments efficiency, and to increase automation. Tagging dNTP's using flurophores[99], or radioactive phosphorous, enables far greater throughput for this method. Labeling the chain terminating didNTPs, rather than dNTP's is also a viable option, and permits Sanger sequencing to be conducted in a single reaction, rather than 4 separate ones. This technique, known as dye-terminator sequencing, utilizes 4 separate fluorescent dyes corresponding to each of the four didNTPs. Following PCR, the fragments are then run through capillary electrophoresis, and stimulated with a laser to cause the terminating nucleotides to fluoresce.

This fluorescence is captured by a fluorescence detector, and can then be translated into sequence data. Numerous high-throughput DNA sequence analyzers have been designed to interpret such fluorescent data, making this method one of the more popular variants of Sanger sequencing. The GSTP1 amplicons produced via PCR, and validated via gel electrophoresis were sent to Genewiz for Sanger sequencing analysis. The consensus identities generated can be seen below:

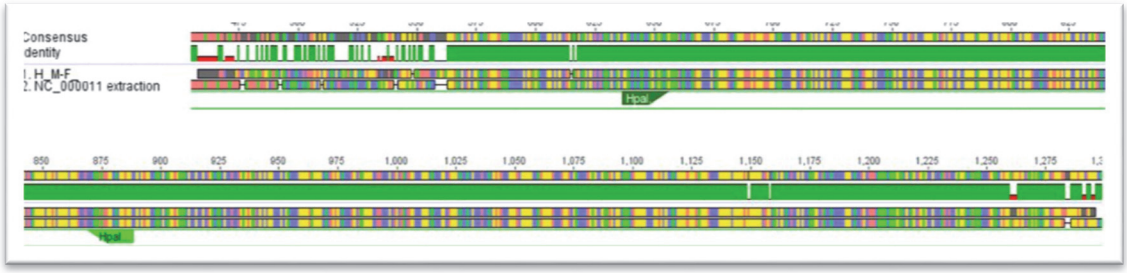


Figure 14: Consensus Identity for H/M_1 Amplicon vs. Genomic Target.
The figure juxtaposes the generated Sanger sequence data with that of the target in the human genome, shown as lines one and two respectively. The green line above the sequences represents their overall agreement, indicating that the amplicon was reproduced successfully. The two green arrows below the sequences represent the location of the designed primers.

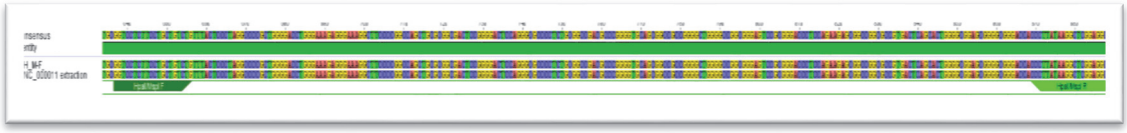


Figure 15: Consensus Identity for H/M_1 Amplicon vs. Genomic Target; HpaII/MspI Primers.
This figure represents the same consensus identity, highlighting the target H/M_1 primers. The result indicates that the generated material is safe to use in later experiments, as this H/M_1 amplicon was reproduced correctly.

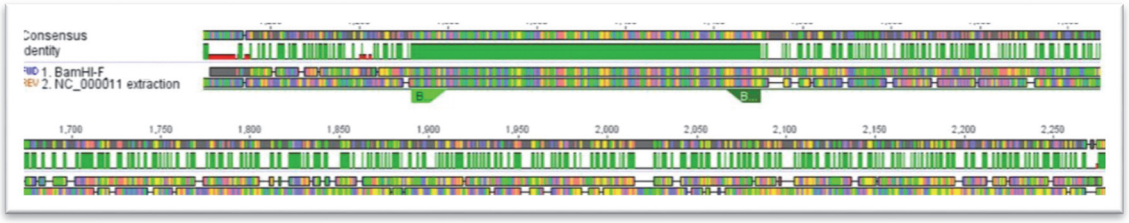


Figure 16: Consensus Identity for B_1 Amplicon vs. Genomic Target.
The figure once again juxtaposes the generated Sanger sequence data with that of the target in the human genome. Compared with the H/M_1 amplicon, there appears to be far less consensus between the B_1 Sanger sequence material, and the target region of the genome. However, the region between the two B_1 primers (shown as two green arrows) exhibits perfect consensus, indicating that the B_1 amplicon was reproduced successfully.

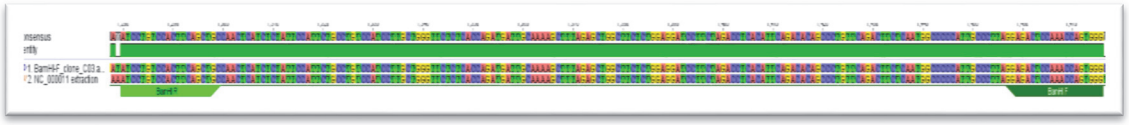


Figure 17: Consensus Identity B_1 Amplicon vs. Genomic Target; BamHI Primers.
This figure represents the same consensus identity, highlighting the target B_1 primers. The result indicates that the generated material is once again safe to use in later experiments, as this B_1 amplicon was reproduced correctly.

Results indicated that all amplicons had been transcribed completely and accurately, an important control before inserting the genetic material into bacterial plasmids, and repeating the process.

3.4 – Molecular Cloning

Molecular cloning is a process used by biologists for over forty years,[100] which isolates specific fragments of DNA, and links them through traditional 3′ → 5′ phosphodiester bonds to a special DNA molecule known as a vector. The significance of such vectors, is that they can be inserted into separate host cells *in vivo*, where they, along with any bound genetic material, can be expressed directly, or replicated along with the host DNA. By binding genetic material of interest to

cloning, or expression vectors, these pieces of recombinant DNA can be easily amplified, sequenced, safely stored for long periods of time, and used for genetic functional studies such as this one.

Two types of vectors are most commonly used for molecular cloning: bacteriophage λ vectors[101], and *E. coli* plasmid vectors, which are ring-shaped extrachromosomal DNA fragments that replicate within a cell. Plasmids are naturally occurring circular dsDNA molecules that are retained separately from a cell's chromosome. This exochromasomal DNA can be thought of as almost parasitic in nature, utilizing cellular machinery to duplicate itself every time the host cell divides. However, many naturally occurring plasmids contain genes which benefit the cell in some way, making this relationship more symbiotic than it would initially seem.[101]

For example, many plasmids encode for the production of enzymes which protect against antibiotics, presenting a major problem for treating bacterial pathogens over time. Additionally, certain plasmids contain so called "transfer genes", which generate proteins which can form into tubes known as "pili", allowing for the transmission of plasmids to other host cells of similar species.[102] This enables drug resistant plasmids to spread even faster, compounding the difficulty faced by modern medicine in preserving the usefulness of antibiotics.

Though these qualities cause difficulties in modern medicine, plasmids prove to be quite useful in the realm of molecular cloning. In

contrast to the naturally occurring plasmids found in *E. coli*, cloning plasmids tend to be roughly 3kb in length, cut down to contain only the genetic material necessary for replication, and an insertion region for the ligation of the DNA of interest, making it far easier to work with. Additionally, the inclusion of various design elements, such as plasmid regions that code for resistance to antibiotics, or toxic minigenes which are disabled through the ligation of new genetic material, are often utilized.To ensure long term transcriptional fidelity of the target amplicons, purified samples were inserted into competent *E. coli,* using pMiniT from NEB as a cloning vector, using the following ratios

	LIGATION REACTION	NEGATIVE CONTROL	POSITIVE CONTROL
Linearized pMiniT Vector (25 ng/µl)	1 µl (25 ng)	1 µl (25 ng)	1 µl (25 ng)
Insert*	1–4 µl*	–	–
Amplicon Cloning Control (1 kb) (15 ng/µl)	–	–	2 µl (30 ng)
H_20	to 5 µl	4 µl	2 µl
Cloning Master Mix (2X)	5 µl	5 µl	5 µl
Total Volume	10 µl	10 µl	10 µl

Figure 18: Recipe for Vector Insertion Protocol
NEB® PCR Cloning Kit Instruction Manual. Ipswich, MA: New England Biolabs, 2012. Web. Apr. 2014.

Reagents were mixed with an appropriate amount of each insert, determined through the following calculation: (3) (25 ng vector) (bp of insert / 2525 bp of vector). Each reaction was incubated at room temperature for 15 minutes, then quickly submerged in ice for another 2 minutes. Simultaneously, a sample of 50 µl NEB 10-beta competent *E.*

coli was removed from storage and allowed to thaw in ice for 10 minutes. Following the completion of both steps, timed to coincide, 2 µl of ligated insert was added to the competent cells, and gently mixed by flicking the tube several times. This combined reaction was then incubated on ice for another 30 minutes, followed by a heat shock of the sample at 42 °C for 30 seconds.

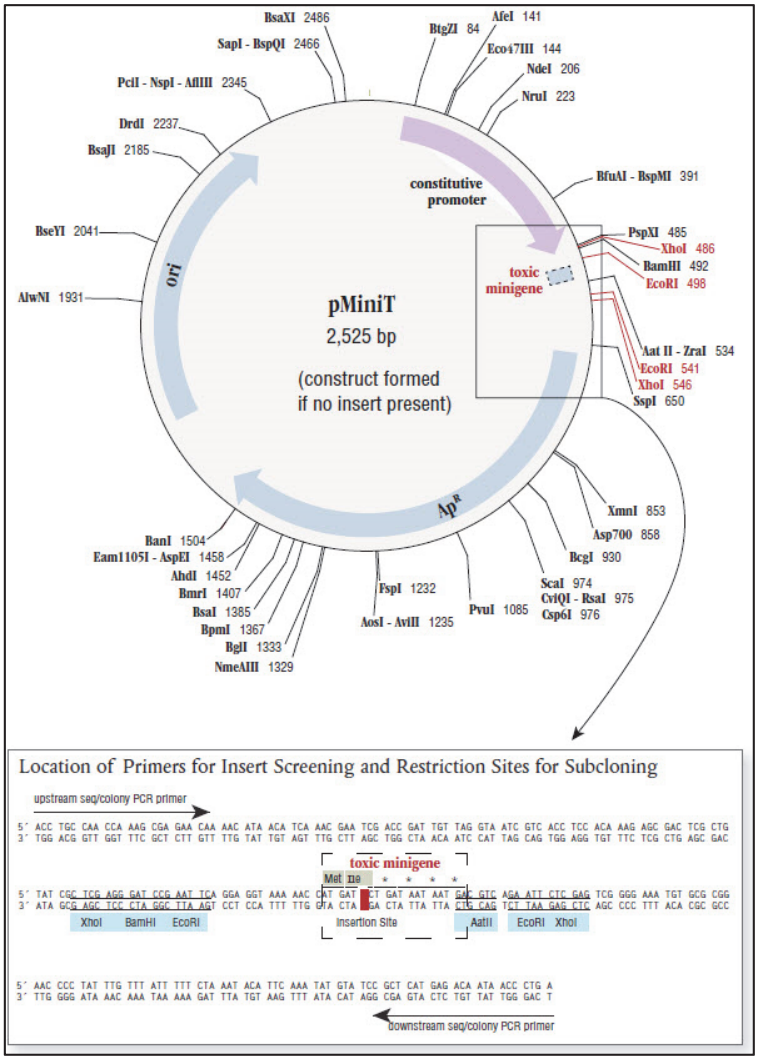
The purpose of this incubation is to allow the bacterial membrane of the competent cells to stabilize, and then become coated with cations released by the ligation enhancers (primarily calcium cations provided from CaCl). This is believed to encourage negatively charged DNA to bind to the phospholipid membrane, though the mechanism is not fully understood. [103] Once bound, the heat shock temporarily alters the fluidity of the membrane of the competent cells, allowing the recombinant DNA to enter at a far more efficient rate. Following heat shock, the mixture is then incubated on ice for another five minutes, to reduce the thermal motion of DNA, ending the transformation step.

950 µl of Super Optimal Broth with Catabolite repression (SOC)[104] outgrowth medium is then added to the mixture, and it is left in a rotating incubator set at 250 rpm, and 37 °C, for one hour. Once completed, the mixture is gently mixed by flicking the side of the tube, and 50 µl of the outgrowth is then spread onto pre-warmed agar plates containing 100 µg/ml ampicillin. The plates are then left to incubate overnight at 37 °C, allowing individual colonies to form. Once growth is

visible to the naked eye, individual colonies can then be carefully harvested using a spatula, and immersed in 5 µl of Lysogeny Broth (LB), also containing 100 µg/ml ampicillin, and left to grow for another 16-18 hours rotating at 250 rpm in a 37 °C incubator. 50 µl of the outgrowth solution is once again spread onto a second set of pre-warmed ampicillin plates, left to grow overnight, producing clonal colonies.

By design, the pMiniT vector contains two key mini-genes, the first being an ampicillin resistance gene, and the second being a toxic gene which inhibits protein synthesis. When genetic material is ligated to this vector, it is inserted directly at the site of the toxic mini-gene, thus nullifying its effect on transformed cells, while preserving ampicillin immunity. Any cells that absorb unmodified pMiniT during transformation will die as a result of the active toxic mini-gene, while cells that had not integrated any vector at all would die when grown out on agar plates containing ampicillin.

Figure 19 pMiniT Insert
NEB® PCR Cloning Kit Instruction Manual.
Ipswich, MA: New England Biolabs, 2012. Web. Apr. 2014.



The only cells remaining would be that population that had integrated the pMiniT vector (containing purified DNA amplicon) into its plasmid DNA. This population would then replicate this inserted recombinant DNA along with its original host genetic material, generating a limitless supply of accurately transcribed target amplicons contained within the plasmid DNA of decedent bacterial clones. A colony of these affected clones can then be stored long-term as a bacterial glycerol stock, which can be used to generate more plasmid DNA without retransforming competent *E. coli* cells.

When desired, the DNA of spliced *E. coli* could be harvested by growing out a colony in lysogeny broth, pelleting the growth in a high speed centrifuge, and then lysing the cells. The plasmid DNA can then be purified out from the mixture of cellular debris, and if desired, and checked again for transcriptional fidelity using the same methods outlined above.

3.5 – Endonucleic Binding

As a preliminary to EMSA, and other examination of protein-DNA binding dynamics, several experiments were run examining buffers that might prevent enzymic cleavage of unmethylated DNA. As mentioned previously, type II restriction endonucleases are only capable of cleaving at their target sites if there are magnesium ions present, which provide hydrogen bonds essential for adopting cleavage conformation. By using an Mg^{2+} free buffer, such as 10mM Tris-HCl 50mM KCl buffer, it was hoped that such activity could be halted, permitting examination of their binding behavior without destroying the target amplicons.

This buffer combination worked exceedingly well, protecting the effected DNA in all experiments as long as a complete absence of Mg^{2+} was maintained. It is worth pointing out here that the presence of methylation can be confirmed visually by once again examining an increased gel shift due to the increased mass.

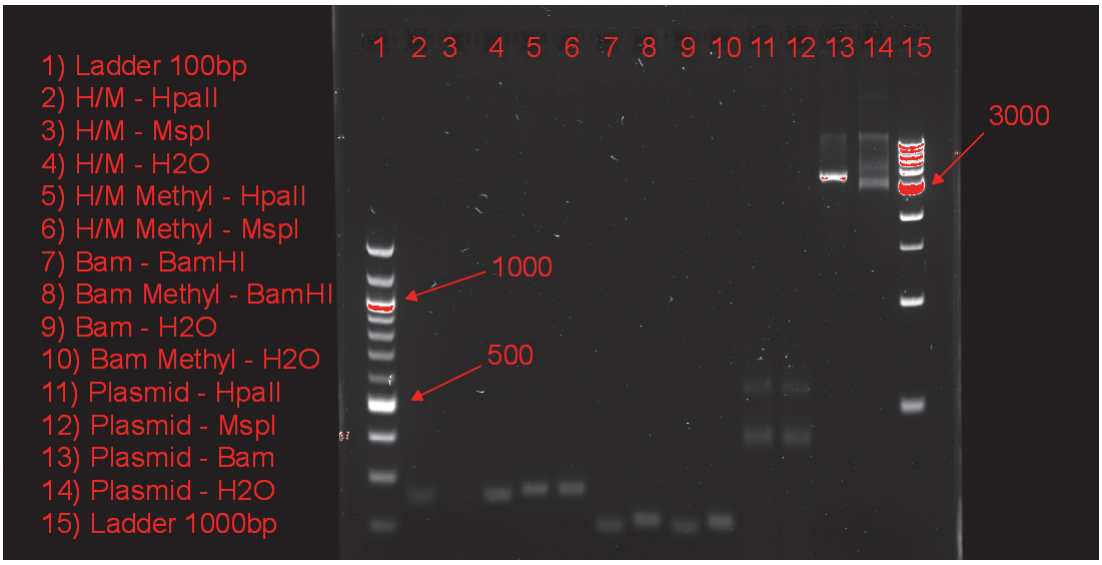


Figure 20: Endonuclease Digestion Gel
Each lane contained DNA sample incubated with an endonuclease, or water as a control. Note that the methylated samples are slightly heavier, and are resistant to cleavage (see lanes 3 and 6). Cloned plasmid matieral was also tested, and was cleaved at the 6 HpaII/MspI target sites, resulting in several smaller segments seen in lanes 11 and 12.

Having confirmed that the amplicons would be protected from cleavage, further data concerning their attractiveness to these endonucleases could be undertaken through EMSA.

3.6 – EMSA

As mentioned in previous sections, EMSA has been used in characterizing DNA binding proteins of all kinds, and has been of key importance in the identification and verification of MBPs. While not designed for high throughput, it is a robust and highly malleable method that can be adapted to accommodate numerous alternate experimental parameters. The amplicons procured in previous steps were designed to contain several target sites for HpaII/MspI, and BamHI activity. By incubating these enzymes with DNA in an Mg^{2+} free buffer solution at a suitable temperature, a sizable amount of enzyme will become bound to

their target strands. Under normal circumstances, the endonucleases would scan DNA strands until finding such a site, then change their binding conformation in order to cleave the strand in two, allowing it to disengage and search for more unmethylated genetic material.

Without the magnesium, these enzymes should remain stuck to DNA, and remain in a protein-DNA complex sufficiently long for examination.Due to the high sensitivity of the assay, only 20 fmoles of each amplicon would be required for each reaction. These were combined with Tris-HCl KCl buffer, along with varying amounts of enzyme and TE buffer on ice, and then incubated at 37 °C for an hour to form protein-DNA complexes, according to the table below.

Table 3: Recipe for Equilibriation Reaction.

Component	**Volume**	**Final Concentration**
Tris-HCl KCl	25 µl	10mM Tris-HCl .5M KCl
DNA	10 µl	20 fmoles
Endonuclease	10 µl	Varied
Tris-EDTA	5 µl	1X
Final	50 µl	N/A

Various neutral osmolytes, such as trimethylglycine were added to these equilibrated samples, as previous studies had shown that their addition had decreased the dissociation rate by a factor of ~75[105,106]. Similar findings had been made concerning the addition of $CaCl_2$, purportedly increasing the half-life of BamHI specific complexes by several hours[107,108], but as this would likely damage the platinum

electrodes in the Lifetech mini-gel tank, this was not attempted. 30 µl of the osmolyte mixture was added in order to “freeze” the reaction state, and prolong endonucleic binding. The concentration of this mixture varied, but proved effective so long as it was at least 200x the molar concentration of the bound DNA.

While this endonucleation step was underway, a 5% 8 x 8 cm polyacrylamide gel was loaded into the tank, submerged in sufficient TBE buffer, and run at 100 V for several minutes in order to draw out any contaminants that might reside in the wells of the gel. Ideally the entire procedure would also take place in a cold room at around 4 °C.

After both steps had been completed, the solutions were then mixed with an appropriate loading dye, and 15 µl of each sample carefully pipetted into the gel. To minimize “dead time”, during which the complexes have yet to enter the gel, electrophoresis was initially conducted at a high voltage (~400 V, or 50 V/cm) for 30 seconds.

Once inside of the gel, as signaled by the loading dye, the voltage is lowered to 80 V, and run for another 45 minutes (10 V/cm). Due to the short half-life of the HpaII and MspI complexes, it was difficult to exceed this time limit without dissociation. Fortunately, 45 minutes was sufficient time for weight-shifted bands to emerge. The gel cassette was then removed from the electrophoretic apparatus, and laid well side up to dry. Using a gel knife, the cassette was carefully cracked open, allowing the gel to be used in the Novex Semi-Dry Blotter.

3.7 – Semi-Dry Electroblotting

The protein-DNA complexes that propagated down the gel need to be transferred to an appropriate membrane for downstream processing. Once accomplished, the bound material can be further analyzed in a variety of different ways, such as using specific antibodies, DNA stains, or ligands. For this method, Zeta-Probe membranes from Bio-Rad were used, though theoretically any appropriate nylon membrane would do. These membranes possess high density positive charge, which strongly attracts negatively charged DNA and protein-DNA complexes. When the gel, the membrane, and several sheets of buffer soaked filter paper are layered in a stack (see below), and are subjected to a strong positive voltage. Given enough time, the genetic material will transfer from the gel to the membrane, and be ready for further processing.

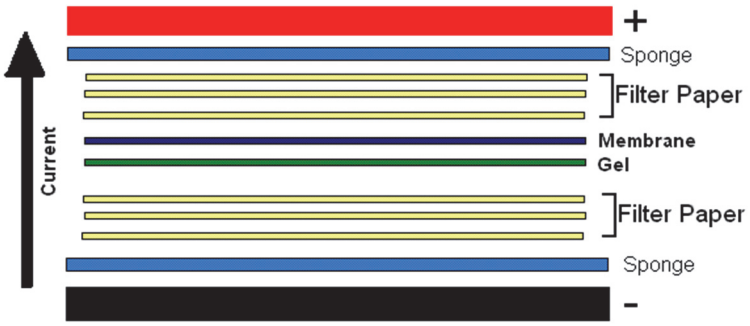


Figure 21: Diagram of Transfer Stack (public domain image)

Two sheets of 2.5 mm blotting paper soaked in TBE were first loaded onto the anode of the apparatus (no sponge is required in semi-dry transfer), followed by the pre-soaked nylon membrane. A roller was used to ensure the membrane remained flush against the filter paper, and removed offending air bubbles. The gel was then placed on top of the

membrane, flattened again with the roller, and then covered up with a final two sheets of blotting paper. The cathode lid of the blotter was then placed on top of the stack, completing the circuit and compressing the components together. The protein-DNA complexes were then transferred at 20 V for 60 minutes at room temperature.

The stack was then disassembled, and the membrane was allowed to dry briefly on some kimwipes, before being transfered to the UV crosslinker. Crosslinking is required to ensure the genetic material remains bound to the membrane as it undergoes further processing. By subjecting the membrane to shortwave (254 nm) UV radiation, the nitrogenous bases of DNA become excited to such a state that they form covalent links with the embedded amine groups. This is easily achieved by subjecting the membrane to a short exposure of UV light, transferring roughly 120 mJ/cm^2.

As mentioned previously, the amplicons designed for this study were produced from biotinylated primers, meaning the forward primer oligos had a biotin modification attached to their 5' end, along with a six carbon length spacer to cut down on steric hindrance.

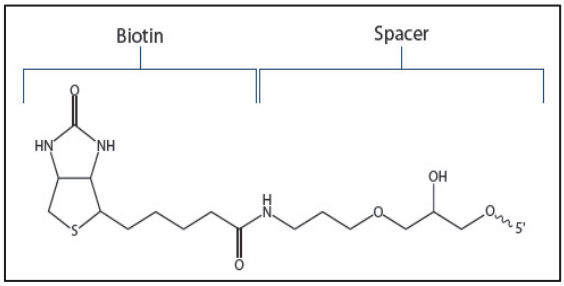


Figure 22: Biotin with Spacer. Digital image. Which Biotin Modification to Use? IDT, n.d. Web. <https://www.idtdna.com/pages/decoded/decoded-articles/core-concepts/decoded/2012/09/20/which-biotin-modification-to-use->.

Biotin is a very useful molecule for a number of molecular biological procedures due to its high affinity for the proteins avidin and streptavidin. By using biotinylated primers to generate amplicons, streptavidin-horseradish peroxidase (HRP) conjugate can bind itself to the resultant DNA. The HRP in turn efficiently catalyzes luminol oxidation, which when combined with certain chemicals, releases high intensity light in a process known as enhanced chemiluminescence (ECL). By immersing the membrane and crosslinked DNA in luminol working solution, the DNA, and protein-DNA complexes can then be imaged and subjected to further analysis.

To bring this about, the membrane was first immersed in 20 mL of blocking buffer for 15 minutes in a shaking container, to prevent non-specific binding downstream. The buffer was then decanted, and then succeeded by 20 mL of another conjugate blocking solution, and once again shaken for 15 minutes. The membrane was then transferred to a new container, and washed in 20 mL of wash buffer for 5 minutes. This step was repeated a total of five times (the buffer being decanted each time), after which the membrane was transferred to yet another new container containing 30 mL of substrate equilibration buffer.

This was left to shake for another five minutes, while simultaneously a working solution was produced by combining 5 mL of Luminol/Enhancer solution and 5 mL of Stable Peroxide solution, avoiding exposure to light. The membrane was removed from the final

round of buffer, blotted on a kimwipe to remove excess liquid, then placed on a clean sheet of plastic wrap, DNA side down. The working luminol peroxide solution was then poured onto the membrane, and left alone for another five minutes.

Once completed, the membrane was briefly blotted on another kimwipe, then wrapped entirely in plastic wrap, avoiding bubbles and folded edges where possible. The treated membrane could finally be imaged for chemiluminescence, allowing for examination of the results of the EMSA experiment.

3.8 – qPCR

A real-time, or quantitative polymerase chain reaction (qPCR) is a technique based off of traditional PCR, discussed above. As in PCR, it uses polymerase and other genetic material to amplify specific target DNA to detectable levels. The process of detection is a separate step in classical PCR, where the amplification product would be run through gel electrophoresis for characterization. In qPCR, this step is avoided entirely, seeing as it utilizes fluorescence detection methods to observe the creation of DNA products in real time as they are produced, and generating quantitative and / or qualitative data. This strategy not only saves time, but has numerous other beneficial aspects, as it removes much contamination risk associated with post-PCR manipulation.[109]

DNA does not fluoresce on its own of course, so certain additional steps must be taken before conducting qPCR. One of the most common

is the inclusion of non-specific fluorescent dye, such as SYBR Green I or EVAGreen, which emit light at 10,000x greater intensity once they are intercalated with double stranded DNA. During PCR, polymerase actively creates more and more double stranded amplicons of the target DNA. As this occurs, the dye binds with these new amplicons, and fluoresces in proportion to the amount of DNA produced in each cycle, and quantifying it in real time.

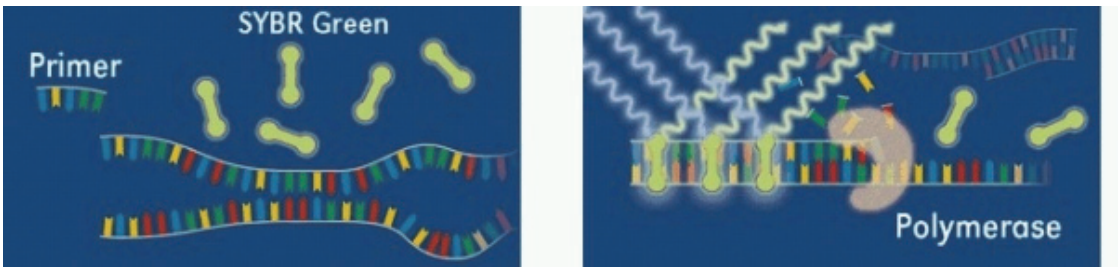


Figure 23: SYBR Green. Digital image.
Genomics & Proteonomics Core Facility. German Cancer Research Center, n.d. Web.
<http://www.dkfz.de/gpcf/lightcycler480.html>.

This particular assay is cheap, and much easier to set up than alternative methods, but it suffers from a number of drawbacks. For one, because of its non-specificity, the dye will bind to any available double-stranded DNA, including unspecific products, primer dimers (forward and reverse primers binding to themselves), and the initial sample DNA, thus introducing bias into the quantification. This can be countered somewhat by running a melting curve after completion of PCR, which can theoretically identify the number different produced DNA amplicons, as well as their relative quantity. Melt curves are not perfect however, so this method tends to be restricted to quantitative analysis of DNA products.

qPCR is also capable of conducting qualitative analysis, detecting the presence or absence of a particular DNA sequence for example. In order to accomplish this, specific detection methods like double-dye probe systems, are required. Originally developed from a similar assay using radio-labeled probes[110], double-dye probes are single stranded DNA molecules with fluorophores attached to their 5' end, and a corresponding quencher attached to their 3' end. The fluorophores emit light as it is excited by the energy of the thermocycler, but has its energy absorbed by the corresponding quencher via fluorescence resonance energy transfer (FRET), so it cannot fluoresce. As these probes are integrated into DNA, polymerase enzymes cut off the 5' end of the probe, releasing the fluorophore into solution, and allowing it to fluoresce freely.

By using several fluorophores, attached to separate sequence-specific double-dye probes, researchers can track the production of several products being created simultaneously, with high accuracy. The method is also highly modular, and can be multiplexed with some forethought. On the other hand, using double-dyes, along with other quantification methods, is more expensive than qualitative qPCR, and tends to be utilized in situations where it is necessary to establish the presence or absence of particular alleles relative to one another.

Integrating qPCR into this method poses several significant advantages. As mentioned above, it can achieve a wide variety of experimental goals at low cost, and can be modified in a variety of ways

to generate qualitative and quantitative data of interest. In addition, it is especially well suited to biophysical modeling, as the plot of fluorescence of a titration standard curve can be used to determine the efficiency of amplification. Such linear regression modeling is not without flaws however, as current methods are highly affected by signal noise, which can produce significant variance in the generated data[111].

For the purpose of validating this method, qPCR was used to probe the integration efficiency of methylated vs. unmethylated dNTPs. As mentioned previously, the generation of fully methylated DNA template had proved difficult to achieve in certain isolated cases, particularly those involving amplifying regions with a high GC percentage. Examining the working efficiency of polymerase working with methylated dNTPs would generate valuable kinematic data. By integrating such findings into currently existing polymerase models, the appropriate ratio of methylated vs unmethylated dNTPs to use in generating randomly methylated amplicons could be determined.

Several preliminary qPCR experiments were run to determine the appropriate quantity of template DNA to use in future experiments, as well as the digestion efficiency of the various endonucleases. The setup of these initial qPCR runs was very similar to that of conventional PCR. DNA amplicons, generated in previous steps, were measured for overall concentration using a nanodrop uv-vis spectrometer. Once measured,

portions of the sample were aliquotted out, and subsequently diluted by factors of 10.

An aliquot with measured concentration of 14.6 ng/µl was utilized for the first digestion experiment. Two separate reactions were generated containing 150 ng of DNA, one intended for digestion, and the other intended as a control, as seen in the following table. Each was produced on ice, and then incubated for 1 hour at 37 °C.

Table 4: qPCR Digestion Recipe

Material	**Digestion**	**Control**
DNA	7.1 µl	7.1 µl
Cutsmart Buffer	5 µl	5 µl
BamHI	1 µl	1 µl
H_2O	36.9 µl	36.9 µl
Total Volume	50 µl	50 µl

Following digestion, the samples were treated using previously described PCR cleanup methods, in order to curtail any further cleavage. The resulting cleaned product was then used as template for qPCR, mixed together with the appropriate forward and reverse primers, and qPCR KAPA SYBR Fast Master Mix (containing ROX as a secondary dye) on ice, according to the following table.

Table 5: qPCR Master Mix Recipe

Material	**Digestion**	**Control**	**NTC**
DNA	21.6 µl	21.6 µl	0 µl
F-Primer	1.44 µl	1.44 µl	1.44 µl
R-Primer	1.44 µl	1.44 µl	1.44 µl
Master Mix	37.44 µl	37.44 µl	37.44 µl
H_2O	10.08 µl	10.08 µl	31.68 µl
Total Volume	72 µl	72 µl	72 µl

Once completed, 20 µl of each mixture were pipetted into three separate wells, making a total of 9 separate reactions. Conducting each reaction in triplicate cuts down on the risk of environmental contamination, or mechanical error adversely affecting the experimental data. The loaded plates would then be inserted into the qPCR thermal cycler, and run according to the programmed thermal recipe. Results and details of subsequent experiments found in the following section.

Chapter 4 – Results:

4.1 – Fully Methylated DNA Amplicons

As mentioned previously, several PCR experiments were run utilizing human male DNA as template, the appropriate primer pairs, and methylated dNTPs. Tests were initially successful in creating a limited set of completely methylated amplicons, specifically those examining the region containing the BamHI cleavage site. However, the majority of regions containing HpaII/MspI targets could not be amplified in the manner previously outlined.

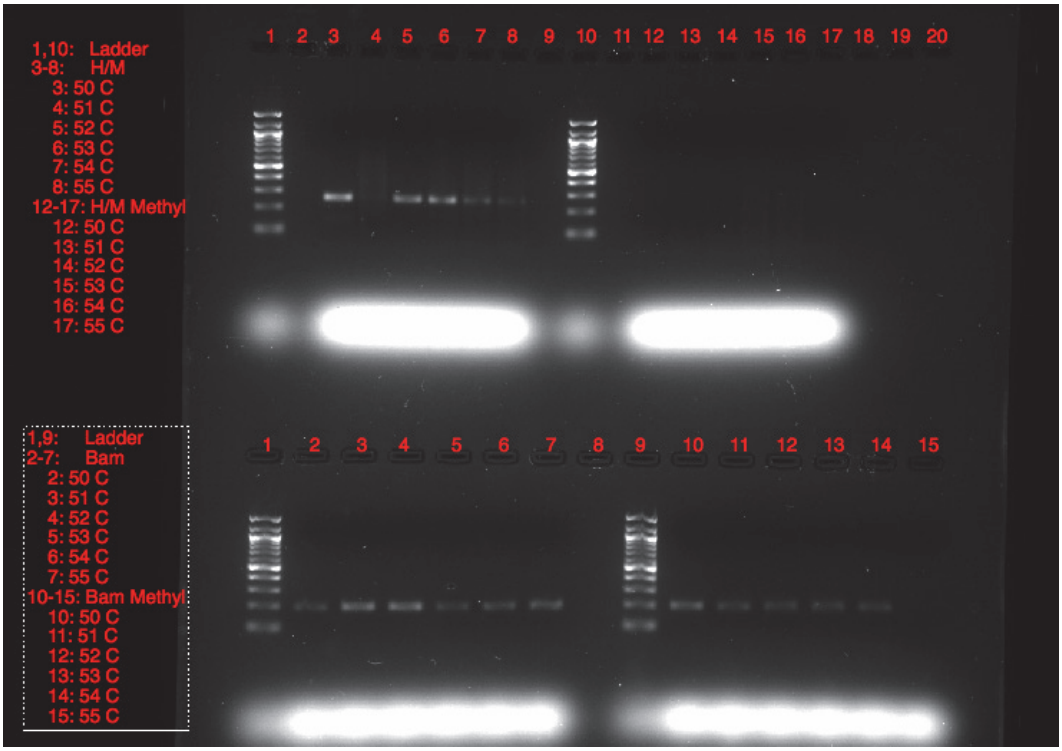


Figure 24: H/M_1 Unmethylated and Methylated Sample, vs B_1 Unmethylated amd Methylated Sample. Note the complete lack of sample for the H/M_1 Methylated sample across all temperatures tested, indicating additional factors were hindering the production of methylated H/M_1.

To resolve this issue, several experiments were run testing the performance of various alternate polymerases, incubation parameters,

and DNA templates. Use of a methylated DNA template, generated using CpG methyltransferase, was also considered, and met with some experimental success. However, the issue was eventually solved through the design of additional primers to be integrated into nested PCR.

That said, further examination of alternate means of offsetting the difficulty of working with 5-methyl-dCTP would be a useful area of future examination, as little work at present has been done on the subject. Considering the relevance of this issue with ongoing characterization of the effects of DNA protein-binding dynamics, new insights into the thermodynamic and biophysical forces at play here would be of considerable interest.

4.1.1 – CpG Methyltransferase

One of the steps taken to ameliorate the issues of working with this difficult set of amplicons, was the usage of CpG methyltransferase (M.SssI). M.SssI is an enzyme capable of methylating cytosine nucleotides it encounters in a CpG doublet. By taking PCR product produced with unmethylated dNTPs, and incubating it with CpG methyltransferase for several hours, all cytosines found in such a CpG configuration will become methylated. As a result of this methylation, the sequence itself will also be unable to be cleaved by methylation-sensitive endonucleases which target CpG doublets, as can be seen below.

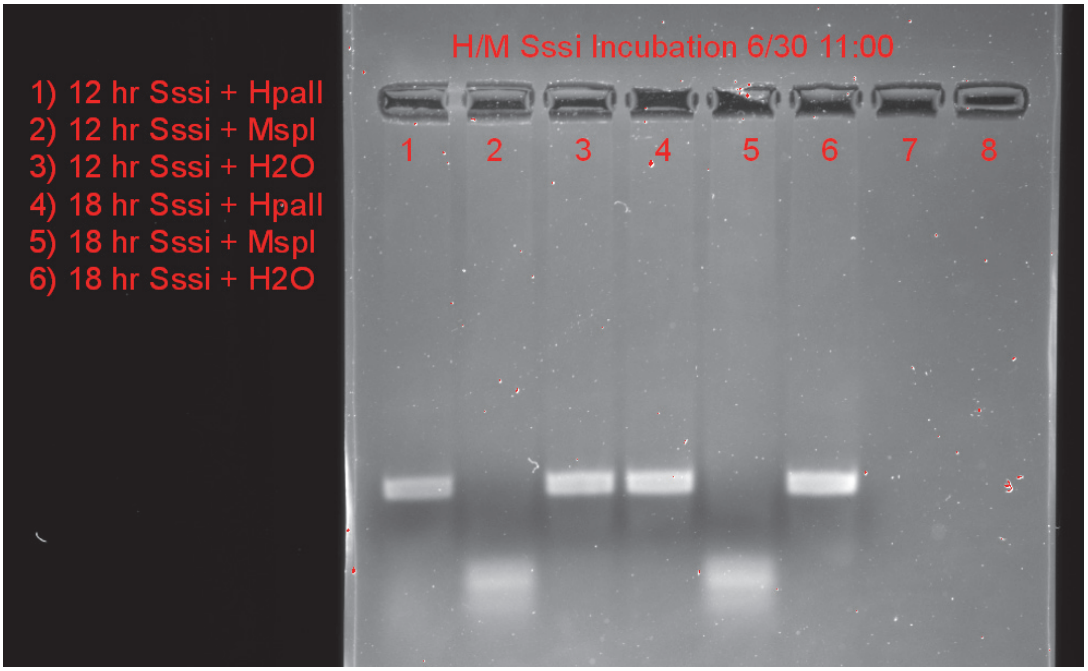


Figure 25: SssI Treated H/M Sample Digestion Gel.
Samples were treated with SssI Methyltransferase for either 12 or 18 hours, and then incubated with an endonuclease, or water as a control. Those samples incubated with MspI appear to be completely digested in both groups, while those incubated with HpaII enzyme emerged unscathed. This indicates the product was successfully methylated in both cases. A a slightly fainter band in lane 1 compared with lane 4 indicated that longer SssI treatment period yielded better results.

By incubating the products with HpaII, MspI, and H_2O as a control, it becomes clear that methylated product has been produced. While MspI is still capable of cleaving the product, methylation sensitive HpaII enzyme is no longer able to attack DNA, and clear bands still emerge in the same position as the control samples incubated with water. Testing suggested that incubations as long as 18 total hours were necessary for complete methylation of all CpG regions of this size, as can be seen in the slightly weaker band for the 12 hour HpaII incubation. Some product was lost as a result of incomplete methylation.

However, such modification methods would not be sufficient for the purposes of this study, as they only result in the methylation of

cytosines in CpG doublets. In the sequence 5' – CCGG - 3' for example, only the second C would be altered by this treatment, leaving the external cytosine unchanged. This is sufficient for the purposes of protecting the sequence from endonucleic cleavage, hence its usefulness for *E. coli,* but not for a method generating entirely random methylation.

It was initially hypothesized that the difficulty of generating completely methylated H/M_1 might be ameliorated by using an alternate DNA template, such as entirely methylated genomic DNA, or DNA partially methylated via CpG methyltransferase. However, as this added additional complication to the method, several additional primers were designed to generate larger amplicons that contained within them either the H/M_1 or B_1 regions of interest, for nested PCR.

4.1.2 – Nested Polymerase Chain Reaction

As its name implies, nested PCR utilizes two sets of primer pairs, where the first pair is used to generate product which will be used as DNA template for the second pair of primers (in a subsequent PCR run). Several of such pairs were designed to nest successful amplicons

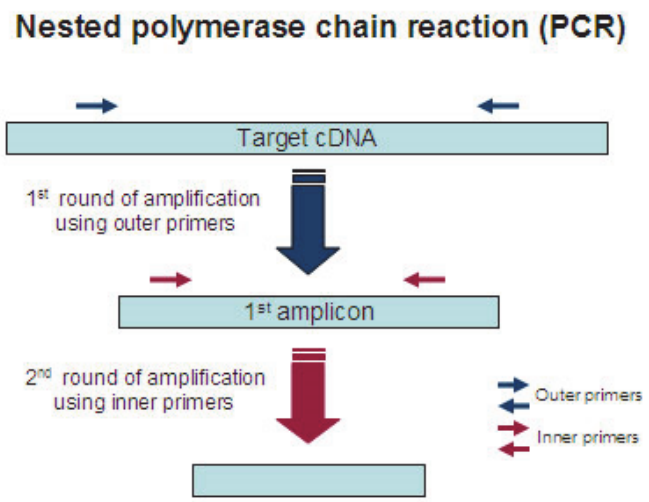


Figure 2614: Nested PCR Diagram. Digital image. N.p., n.d. Web. <http://invitrosperm.blogspot.com/2013/12/lavado-seminal-como-alternativa-de.html>.

within larger regions of roughly 1000-2000 base pairs in length. However, the majority of those designed performed rather poorly, likely due to the high GC content of the region, despite several attempts to optimize thermal recipes for each product.

Eventually, two appropriate pairs were identified that amplified sufficiently well to be used for nested PCR with H/M_1 and B_1, namely H/M_61C, and B_1.1K. These samples were generated in the same manner previously described, and can be seen below. H/M_61C was the strongest candidate by far with high yield and consistency, though only producing product when the annealing temperature of the thermal recipe was set above 63 ° C. All three tested extended amplicons for B_1 were successful, though 1.1K had the best overall performance and yield.

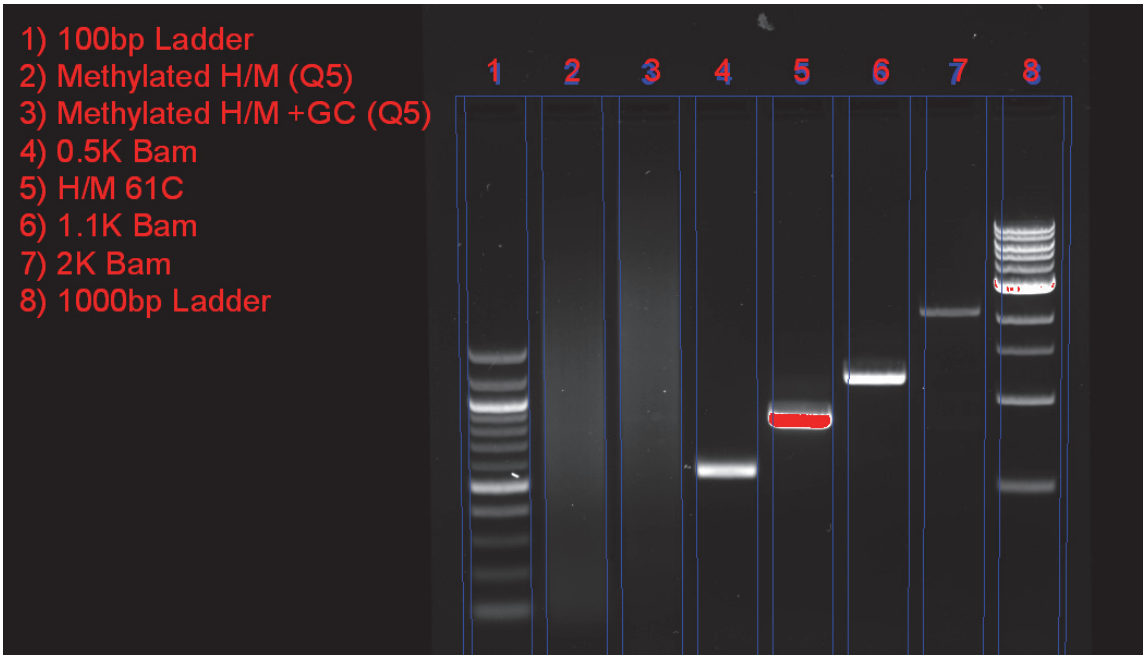


Figure 27: Gel Electrophoresis of External Nested Amplicons.
Lanes 4 through 7 contain sample of amplicons containing either H/M_1 or B_1 nested within them. Lanes 2 and 3 were a failed attempt to use an alternate polymerase (Q5), and show clear signs of streaking. All other lanes were successful, and presented bands where expected.

These amplicons were then purified, and used in successive rounds of PCR with the original primer sets, and methylated dNTPs. This at last produced the appropriate methylated products, as can be seen in the gels below.

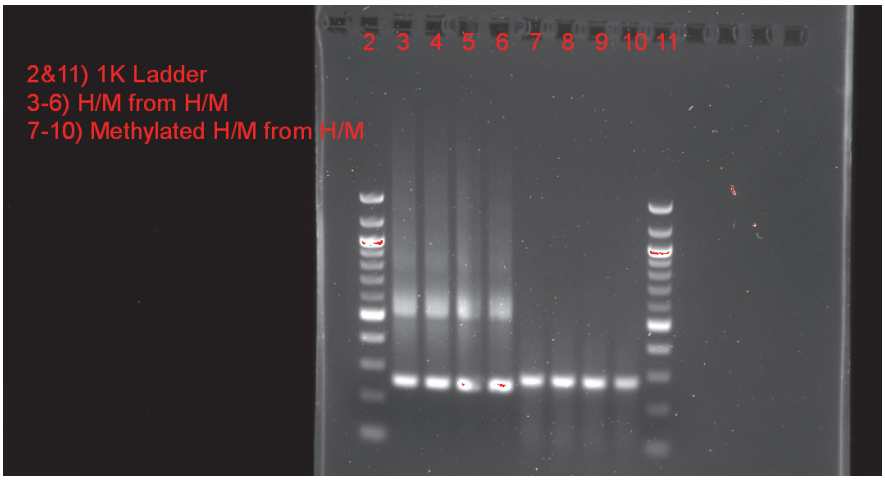


Figure 28: Methylated H/M_1 Generated from Nested Material.
H/M_1 amplicon was generated via nested PCR reaction, using either H61C or H/M_1 amplicon as a DNA template for amplification. Lanes 2 through 5 used H61C and normal dNTP's performed poorly, presenting with a large secondary band due to an overload of template DNA.
Those using methylated dNTP's used H/M_1 amplicon as template, and appeared to be generated successfully. Note the slight weight shift due to added molecular weight from methyl groups, indicating successful generation of methylated product.

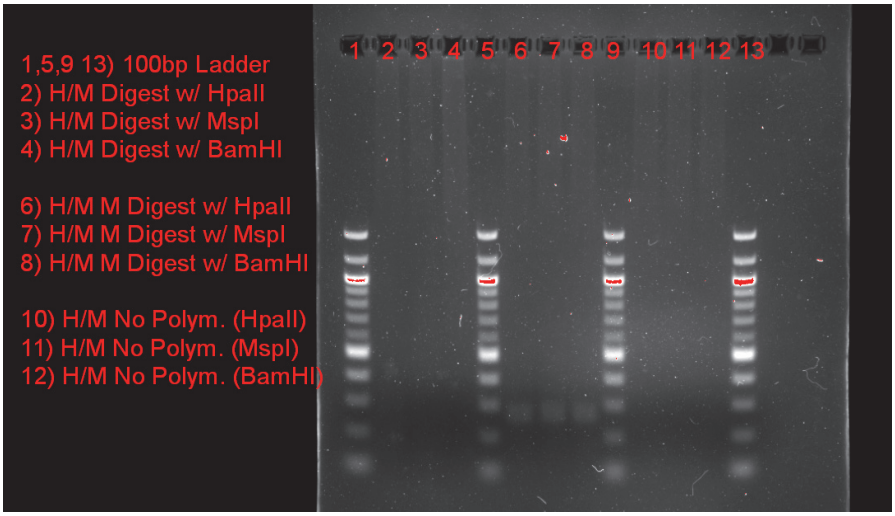


Figure 159: Verification of H/M_1 Product Methylation.
Samples of methylated and unmethylated H/M_1 were generated, and then digested with one of three endonucleases. Unmethylated sample was digested in all three cases (even by BamHI, indicating star activity), while methylated H/M_1 was unscathed. A secondary control was run in lanes 10 through 12, generating H/M_1 amplicon, using H/M_1 as a template, with no polymerase. Bands present where expected, indicating that the central samples were completely methylated.

Though subtle, it can be noted even in this image that the procedure has been a success. The slight upward shift of the methylated H/M_1 amplicons seen on the right can be attributed to the additional methyl groups found on each cytosine. Methylation status was further verified by digesting the material using HpaII, MspI, and BamHI as a control. With this completed, probing into protein DNA interactions could begin.

4.2 – *EMSA Experiments*

Numerous EMSA experiments were conducted in order to optimize protein-DNA complex half-life, and transfer fidelity. One of the first of such experiments was conducted to determine the sensitivity of the assay itself with our imaging tools. Various 3-fold dilutions of B_1 control amplicon were incubated without endonuclease and run through the described EMSA protocol, producing the image seen below.

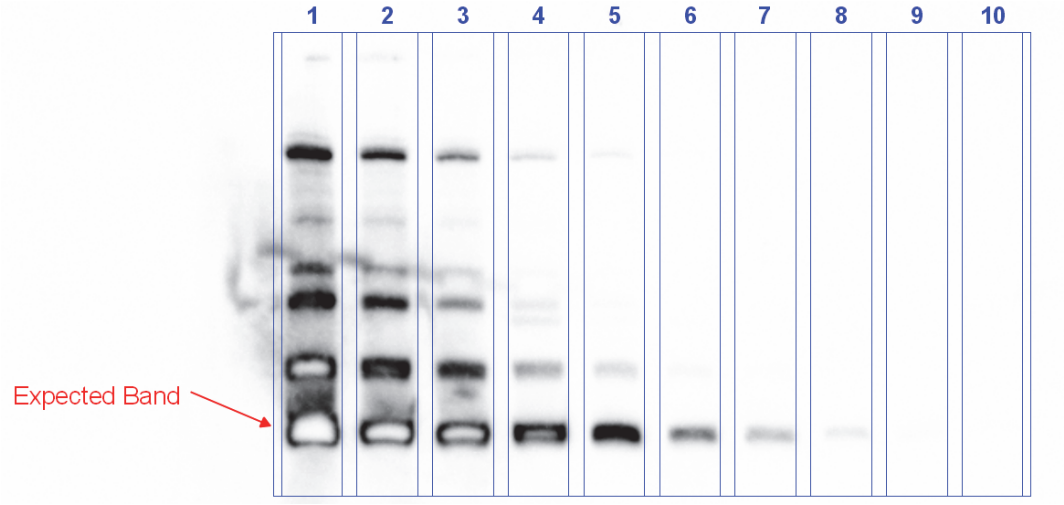


Figure 30: Initial EMSA Experiment.
Numerous extraneous bands indicate insufficiently clean product. The expected band is indicated by the red arrow to the left. As the samples become more dilute, the extraneous bands fade away more rapidly than those of the target bands, indicating they represent the majority of sample.

Table 6: Amount of DNA Used in Initial EMSA Experiment

Lane No.	1	2	3	4	5	6	7	8	9	10
fMoles B_1	2,974	991	330	110	37	12	4	1.4	0.5	0.2

The assay itself was more sensitive than initially expected, and revealed the considerable impurity of the sample utilized. The lowest band was the anticipated B_1 amplicon, but the majority of wells presented with additional bands above this point at lower intensity, indicating contamination of the product. Later examination of the sample traced this back to non-specific priming in PCR, which was resolved by increasing the annealing time and temperature. The results of the modified procedure can be seen below.

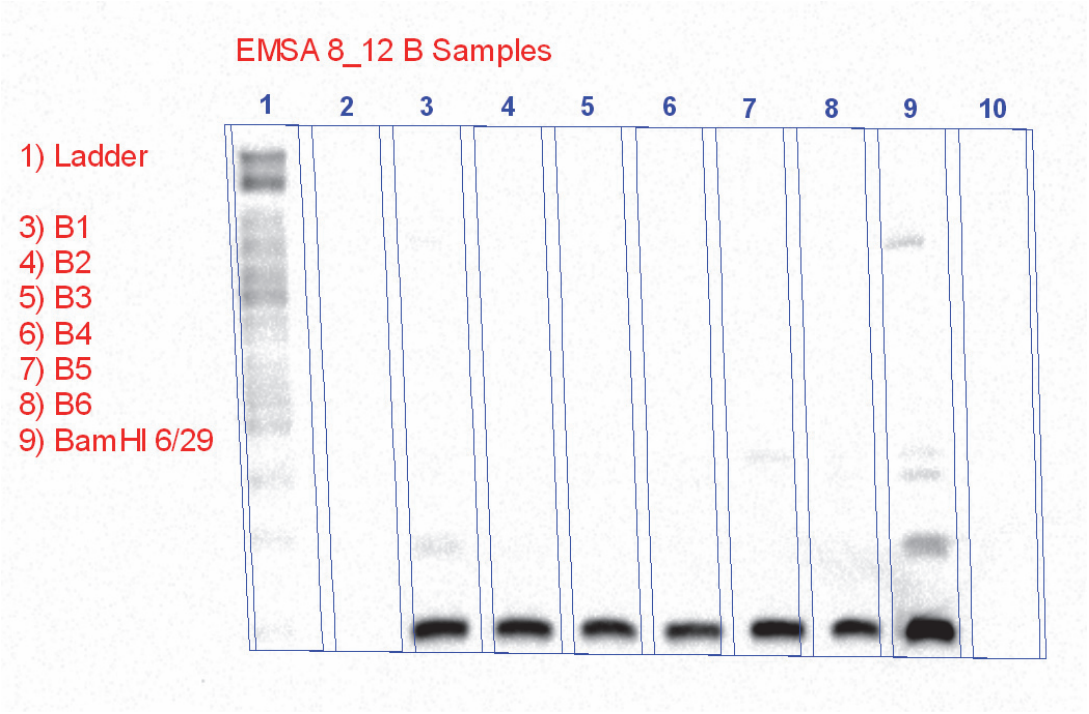


Figure 31: Modified EMSA Experiment. Bands uniform and presented where expected in gel. Old sample shown in row 9 for comparison. Two of the new samples tested, B1 and B5 still presented with extraneous bands. All other samples were sufficiently clean.

Using the appropriate amount of clean DNA results in clear bands presenting where expected on the gel. For comparison, the original sample used in the previous image can be seen in well 9. A DNA ladder was also utilized for this experiment, though it presented fainter bands than were hoped. Initial tests intended to preserve bound protein-DNA complexes were not very promising. No matter how much enzyme was used, or how rapidly EMSA could be completed, nearly all bound endonuclease appeared to separate from its target DNA by the time it had been transferred to the nylon membrane, resulting in uniform bands without weight shift, as seen in the image below.

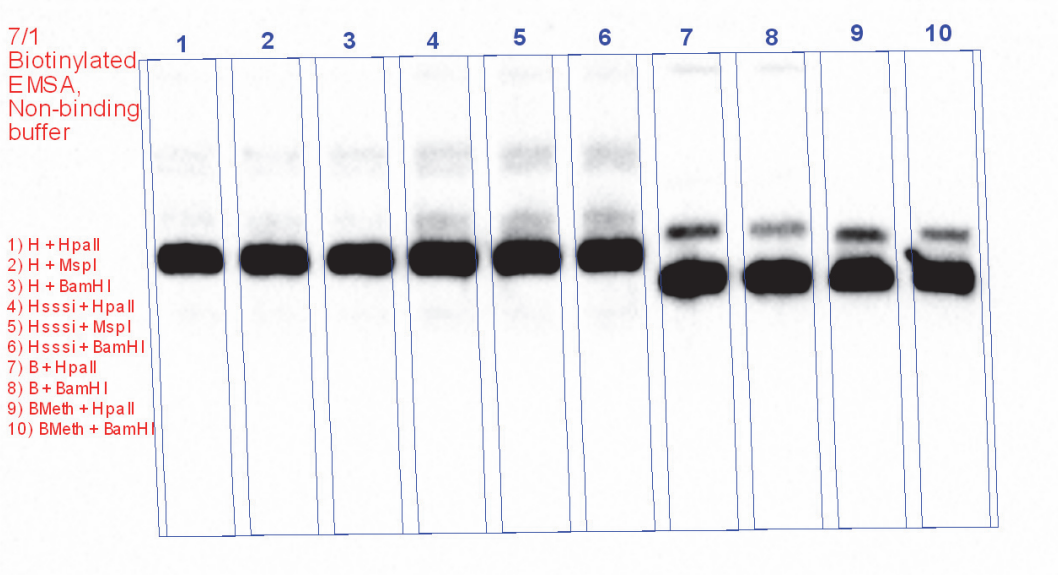


Figure 32: Negative EMSA Result
While DNA samples were successfully protected from cleavage while incubated with various endonucleases, none of the lanes presented with a strong, discernable weight shift; indicating that any protein-DNA complexes had since dissociated. Only a small amount of DNA-protein complexes remained bound, as evinced by faint trailing bands.

The addition of neutral osmolytes as a means of extending complex half-life, as well as utilizing concentrated endonuclease, gave promising

results. B_1 amplicons were designed to have only one BamHI restriction site, meaning that if BamHI were to bind to its target site and remain as a stable complex, a distinct upward shift should be observed. While a shift was evident, indicating that some amount of endonuclease had bound to the available DNA, the bands presented were smeared, indicating that rapid association and dissociation of the complex was occurring. All else being equal, the more available enzyme in the local vicinity, the more likely a given restriction site is occupied at any given time. The greater the amount of restriction enzyme therefore, the greater the shift, until a point where there is sufficient enzyme to make all available DNA functionally permanently bound.

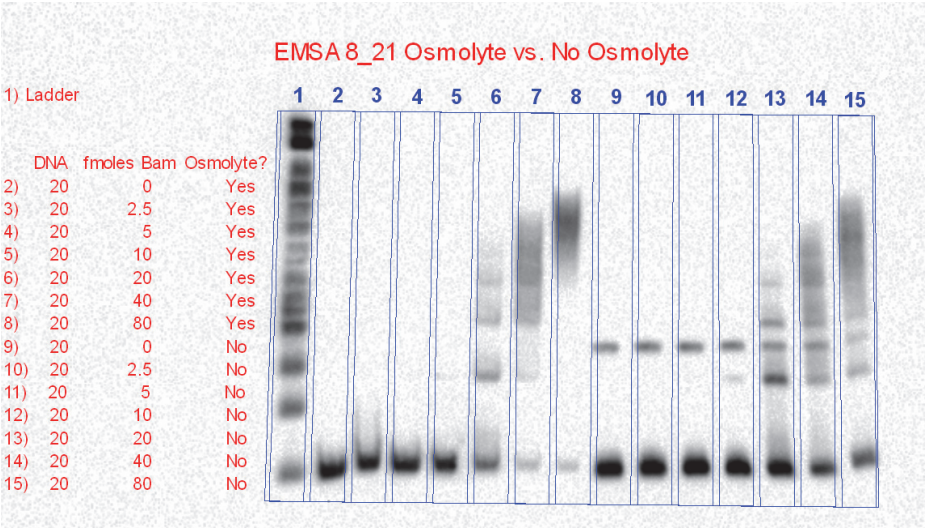


Figure 33: Neutral Osmolyte Experiment.
The lanes on the left were treated with neutral osmolytes, while those on the right were not. At lower concentrations of BamHI, there is little sign of weight shift in either set of samples. However as concentration increases, this shift becomes far more noticeable and uniform in the osmolyte treated samples, for example in lanes 7 and 8. An additional band seen in the osmolyte-free samples can be seen, but was determined to be a result of sample contamination, and not of bound protein-DNA complexes as they appear in lanes with no enzyme at all.

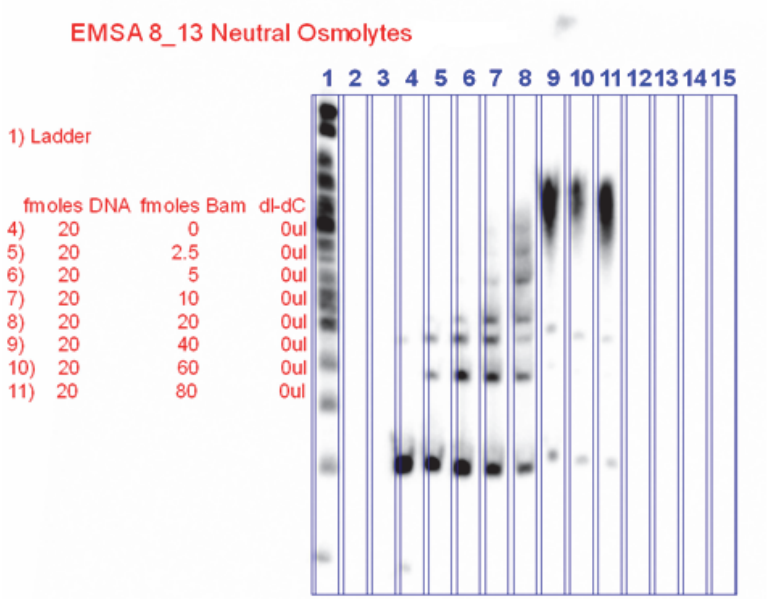


Figure 34: Additional EMSA Experiment with Osmolytes
Repeat of previous test. Note again the signs of streaking seen in lanes 9 through 11. These indicate a complete saturation of all binding sites on the DNA sample, with a weight shift corresponding to the additional BamHI enzyme.

This last point can be illustrated quite well in the following image. Utilizing neutral osmolytes, as well as other optimization techniques, clear bands emerge when the molar concentration of BamHI exceeds that of the DNA strands by a factor of over 200.

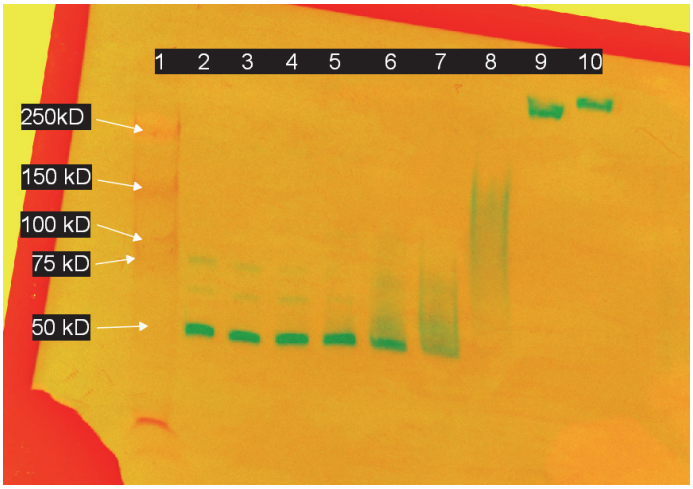


Figure 35: Effect of Osmolytes on B_1 Protein Complex Half-Life.
This run used non-biotinylated ladder that presented visual bands, as opposed to fluorescent ones. The image is a composite of two separate imaging techniques, colorized for ease of visualization. Due to far higher concentrations of enzyme, lanes 9 and 10 show signs of complete saturation.

Table 7: Amount of DNA and Protein Used in Osmolyte Experiment

Lane No.	**fMoles BamHI**	**fMoles B_1 Sample**	**BamHI/B_1 Ratio**
1(Ladder)	0	0	N/A
2	0	20	N/A
3	6.67	20	0.33 : 1
4	20	20	1:1
5	60	20	3:1
6	180	20	9:1
7	540	20	27:1
8	1,620	20	81:1
9	4,860	20	243 : 1
10	12,660	20	632 : 1

This indicates that the method is capable of generating stable protein-DNA complexes of a sort by brute force. The shift itself is far greater than expected, with the molecular weight of a BamHI dimer being roughly 75 kD (kilo Daltons), and the bands in wells 7 & 8 exceeding 250 kD. This would imply that their stability is a result of both specific, and non-specific binding to the strand. Type II restriction endonucleases first bind non-specifically to DNA, then locate their target sites via linear diffusion.[112] Several BamHI dimers therefore could exist on the strand simultaneously, stuck in a traffic jam of sorts, until they themselves dissociate.

The desired shift can be observed in well 6, with an observable band dwelling in the 125-150 kD range, indicating the presence of a

single bound BamHI dimer. This affirmative result however, is hardly ideal. The band is highly smeared, rather than a solid band, indicating that a sizable portion of the complexes had dissociated before the completion of EMSA. Numerous subsequent runs presented similar bands, which appear to be symptomatic of the short half-life of non-cleaving endonucleic activity.

Future experiments will explore potential ways to extend this time period, or to optimize the EMSA experimental protocol sufficiently to render it useful for even the most difficult of protein-DNA complexes. The utilization of formaldehyde crosslinking, to permanently bind endonucleases to DNA after equilibriation, might be a viable first step, though such measures would contaminate data concerning binding strength. Further attempts will undoubtedly shed light on these issues.

4.3 – qPCR Analysis

The results from the initial experiment outlined in the qPCR methods section can be seen in the amplification plot below.

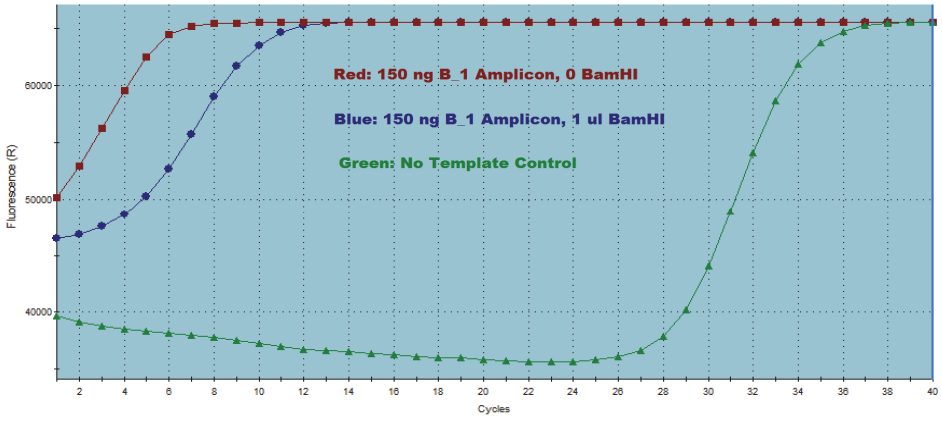


Figure 36m: Amplification Plot of Initial qPCR Experiment. The red and blue lines represent the measured fluorescence of two separate reactions, utilizing 150 ng of B_1 amplicon as starting material, and either 0 or 1 µl of BamHI enzyme. The green line represents a control containing no DNA template at all.

In this and all subsiquent figures the y-axis corresponds to the level of fluorescence detected during amplification after a given number of cycles of amplification, seen on the x-axis. The lines themselves represent the average fluorescence data detected for the three triplicate wells for each reaction, as described above. As more and more amplification cycles takes place, the measured fluorescence of any well containing amplified genetic material will increase, until reaching a saturation point.

The results of this particular experiment indicated that digestion efficiency of these restriction endonucleases had been overestimated somewhat, and that either too little enzyme had been used, or far too much DNA. Despite the fact that an hour of digestion had taken place, there was noticeable fluorescence observed from the digested wells starting at cycle 0 (initial conditions). This is to be expected of course, as the digestion of DNA results in the creation of smaller fragments of that same DNA, rather than its complete destruction. However, this fluorescence was not expected to increase, as the primers and polymerase should be unable to amplify such cut up material.

It seemed very unlikely that contamination of some sort had taken place, as the no template control had become amplified after over 30 cycles of PCR. This, along with the very quick amplification of the undigested material, indicated that far less DNA template would be required in subsequent experiments. To bring this about, a simple

titration run was conducted, examining the PCR amplification curves of various dilutions of the same genetic material. The results of this run can be seen below.

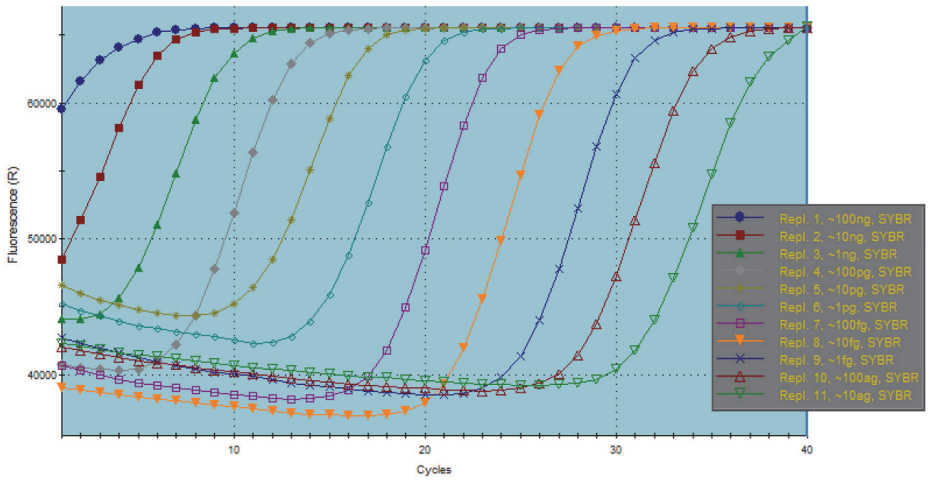


Figure 37: Amplification Plot of B_1 qPCR Titraion
Samples were created via serial dilutions by a factor of 10, and through a standard qPCR reaction. All else being equal, the relative speed at which a given sample reaches fluorescence saturation is dependent entirely on the amount of starting material in each sample reaction. The effect of different quantities of DNA template can be clearly seen in the uniform spacing of each amplification curve.

The results were precisely as expected, with the curves evenly spaced relative to their starting concentrations. Similar curve spread was observed in subsequent tests involving alternate amplicons, as can be seen below.

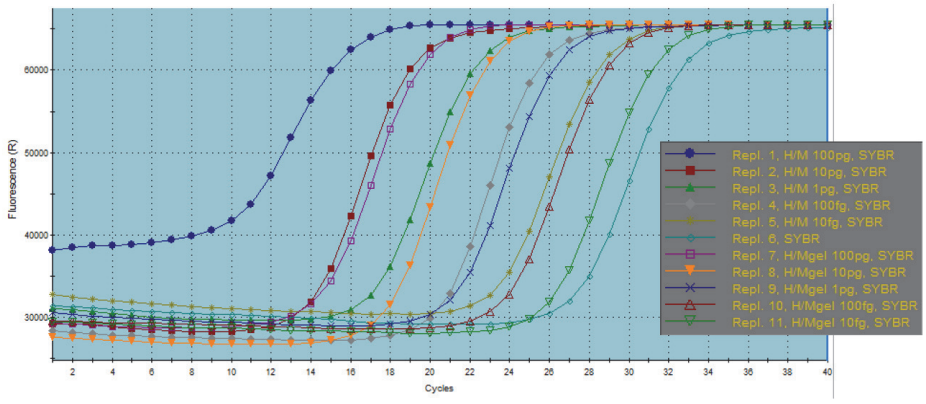


Figure 38: Amplification Plot of H/M_1 qPCR Titration
This run was the result of a similar series of serial dilutions, using H/M_1 samples instead of B_1. These performed as expected, with uniform spacing. Samples 7 through 11 were additional H/M_1 samples that had been purified via gel extraction. These samples were found to be of poor quality.

Given the exponential growth of product during PCR, and the linear relationship between the amount of this product and measured fluorescence, the starting amount of product is the prime deciding factor determining the number of cycles required to achieve fluorescence saturation. With this in mind, a far smaller quantity of starting product was utilized in future trials, hovering around the 1 pg range so as to center the control curve at twenty PCR cycles.

In order to test the rate of integration of varying levels of methylated dNTPs via qPCR, new reaction mixtures would need to be produced containing these altered nucleotides. Proprietary information regarding the KAPA qPCR master mix utilized in previous experiments is not available information, but the key ingredients were known to be SYBR Green, OneTaq polymerase, and 2.5 mM $MgCl_2$[113].

Initial experiments were run utilizing NEB OneTaq, and EvaGreen dye, which has been shown to outperform SYBR Green in terms of the robustness of its generated PCR signal, and its sharper and stronger melt peaks[113]. Using otherwise identical recipes as detailed in the PCR section, with either completely methylated or completely unmethylated dNTPs and 1X concentrated Evagreen dye, several qPCR tests were run to see their relative amplification performance. The amplification and dissociation curves can be seen below.

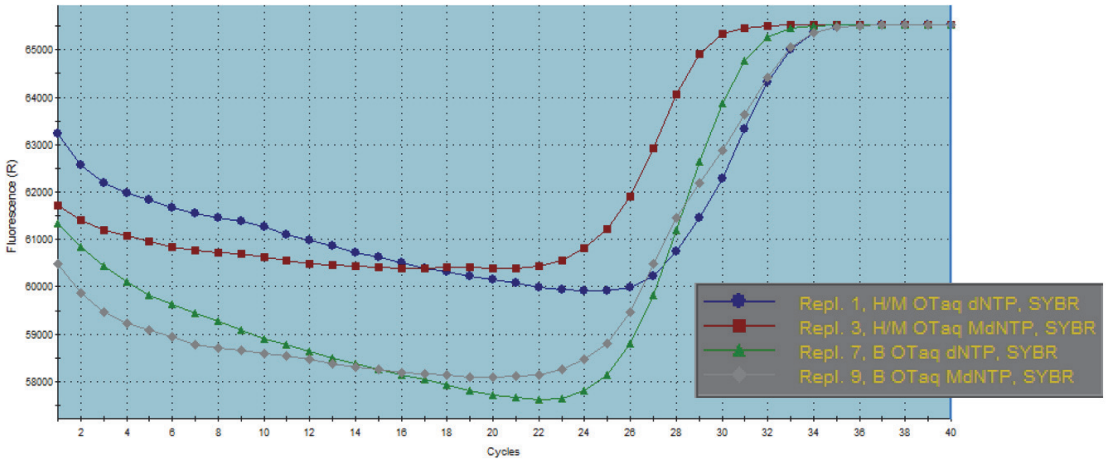


Figure 39: Amplification Plot for Initial EvaGreen qPCR Experiment
Each reaction was generated using the standard recipe, using NEB OneTaq and Evagreen dye. Despite using a relatively high amount of starting material, each reaction reached a saturation point after the expected amount of cycles, regardless of methylation status.

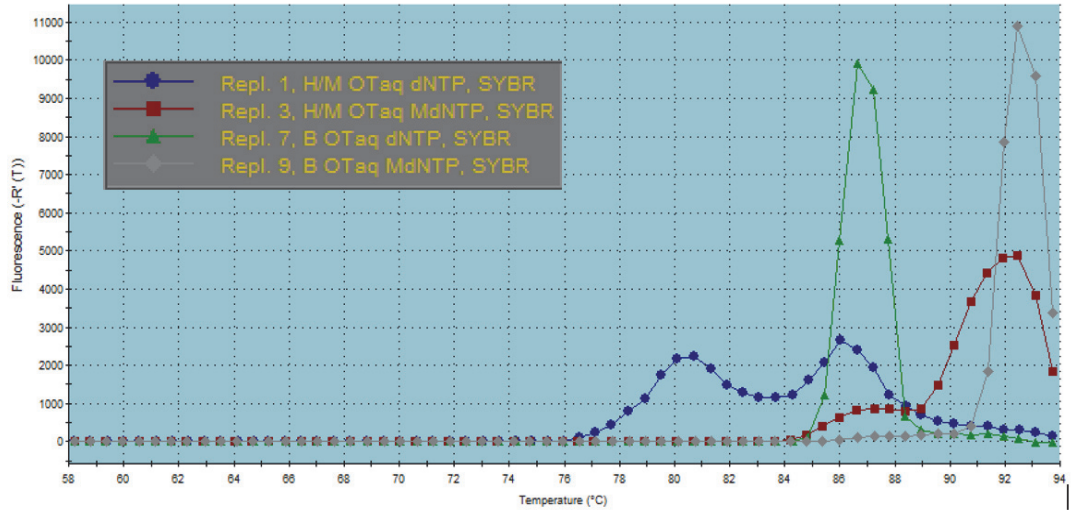


Figure 40: Melting Curve Plot for Initial EvaGreen qPCR Experiment
Melting curves are generated by measuring total fluorescence of sample over a wide range of temperatures. As temperature increases, dsDNA begins to dissociate, freeing the fluorescent dye, and thus lowering fluorescence. Assuming the majority of dsDNA is represented by the target amplicon, the total amount of fluorescence will dip sharply when it's specific melting point is reached. By taking the first derivative of this fluorescence data vs temperature, a peak should emerge indicating the presence of these DNA compounds, centered roughly at the same point. An example of these peaks can be seen here. Note the effect of methylation on peaks for both H/M_1 and B_1, shifting both up several degrees.

Results were mixed, but still largely positive. While each reaction took several cycles longer to reach fluorescent saturation than those using a similar amount of template in the KAPA master mix trials, they

still outperformed the NTC. Additionally, due to the lack of optimization with this novel reaction mixture, sluggish reactivity could be reasonably explained. Examination of the melt curves was far less encouraging however. While the B_1 samples presented with distinct peaks, with the methylated amplicon melting at a slightly higher temperature as expected, the H/M samples appear to have melted with two separate peaks, indicating non-specific, or inaccurate PCR amplification. These findings were cross referenced using programs like UMelt from the Wittwer Lab at the University of Utah. Using various peer-reviewed algorithms, it can produce highly accurate predictions of DNA melting curves and denaturation profiles[114]. The predicted melting curves for H/M_1 and B_1 are shown below.

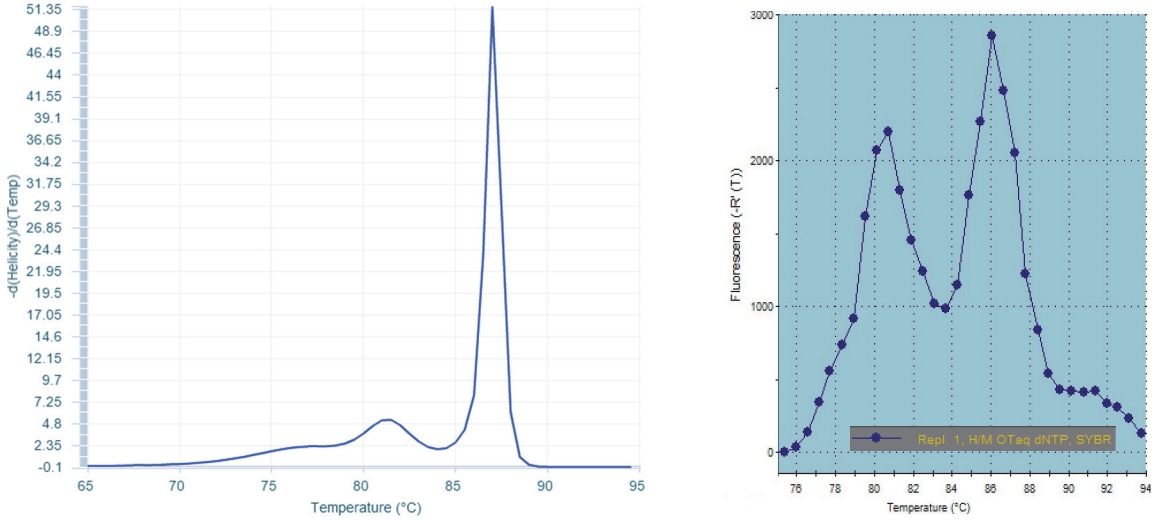


Figure 41: uMelt Prediction for Initial H/M_1 Sample; Compared with Experimental Result. Both images show the derivative of fluorescence over temperature, and the Y axis of the uMelt image is shown in terms of the derivative value. The Y axis of the experimental image is in terms of raw fluorescence, but represents the same curve none the less.
While the experimental curve presents with two peaks corresponding roughly to those predicted by uMelt, the low fluorescence of the secondary peak indicates very poor yield. Further optimization of the PCR recipe was required for H/M_1 samples..

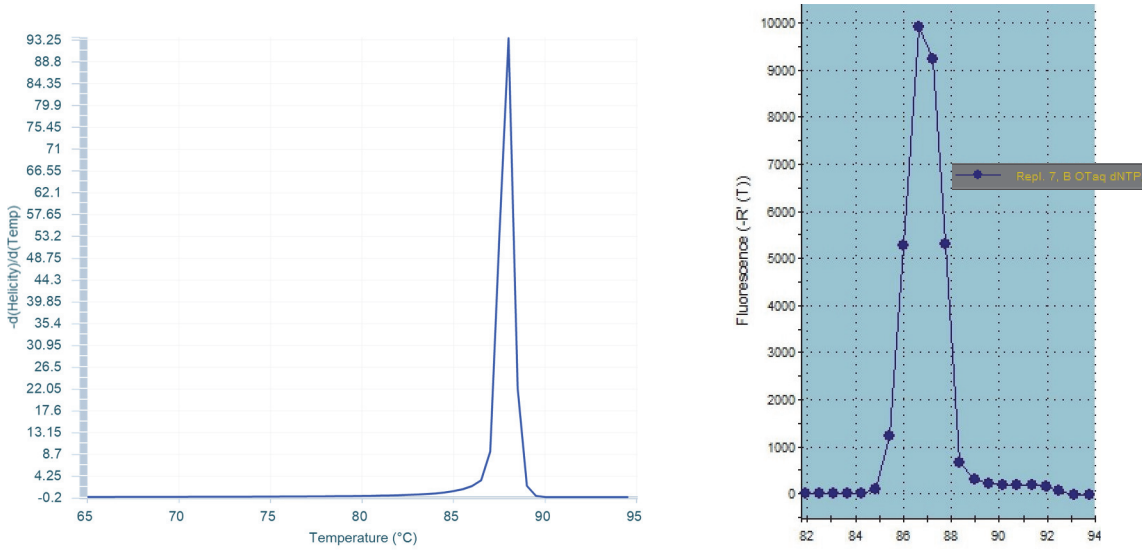


Figure 42: uMelt Prediction for B_1 Sample; Compared with Experimental Result. Both images show the derivative of fluorescence over temperature, and the Y axis of the uMelt image is shown in terms of the derivative value. The predicted and experimental results for B_1 appear to be slightly more accurate, with the fluorescent peak appearing at 87°C.

Even allowing for the slightly more rounded curve, a result of the exceedingly high GC % of H/M_1, the melting profiles of both the methylated and unmethylated samples bore little resemblance to this prediction. B_1 samples on the other hand, matched in both the location of their melting peaks, as well as the general shape of their dissociation curves.

Several attempts were made to optimize the H/M_1 qPCR reaction, with trials of various different polymerases, and alternate thermal profiles being considered. Ultimately, a high increase in Tm, while maintaining the usage of OneTaq polymerase, produced the most promising results. The dissociation curve of this run can be seen below.

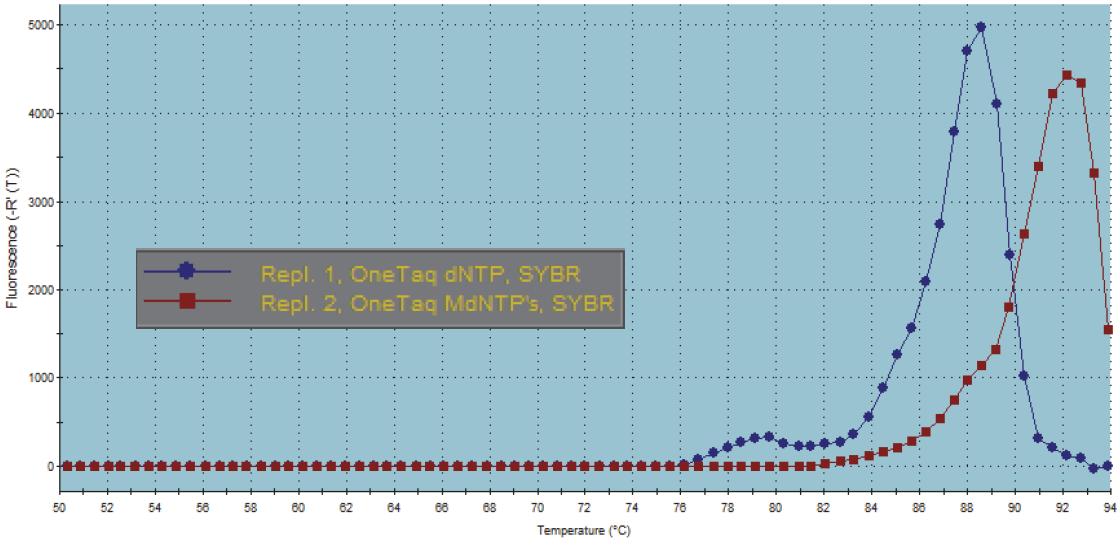


Figure 43: Melting Curve Plot for High Tm qPCR.
Performance for H/M_1 product was improved substantially as compared with the previous graph. Note the uniform shape and high measured fluorescence for both methylated and unmethylated product. Again, the methylated sample melted at an inflated temperature.

The peaks themselves were far more in line with those predicted by uMelt, and were observed in subsequent trials as well. The shift to the right of the graph observed for the methylated dNTP sample can once again be attributed to the presence of methylation, which has been reported to increase overall melting temperature in CpG methylated DNA[115].

The melting curve also appears to have substantially improved, eliminating additional peaks and exhibiting the expected shape and curvature as predicted by uMelt. The new comparison can be seen below.

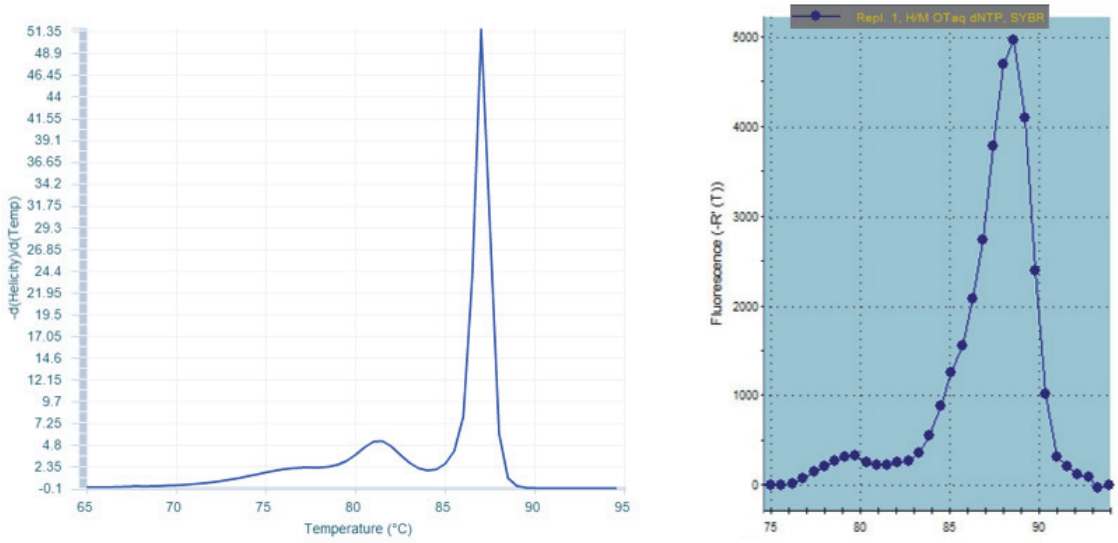


Figure 44: uMelt Prediction for Alternate H/M_1 Sample; Compared with Experimental Result. Both images show the derivative of fluorescence over temperature, and the Y axis of the uMelt image is shown in terms of the derivative value. The Y axis of the experimental image is in terms of raw fluorescence, but represents the same curve, at a different relative scale. While the experimental curve appears to peak slightly higher than predicted, the overall shape of the curve is maintained. Note the smaller secondary peak at 80°C.

However, these results are not proof positive of the presence of completely methylated product, as those peaks might result from non-specific amplification. Further follow up experiments examining these results will be required moving forward.

4.4 – Methylated Libraries

With these positive results, the final step required for evaluation of this method was the generation of methylated DNA libraries. Their creation involved several previously discussed steps, beginning with the amplification of template DNA utilizing varying levels of methylated dNTPs via PCR, followed by a process bisulfite conversion for downstream methylation analysis.

These treated products were then prepped for library integration by ligating NEB Next adaptors to either end of the double stranded amplicons, which serve as platforms for the annealing of specific identifiable DNA barcodes, and new primers for next generation sequencing. Due to the specific design of the amplicons utilized, fragmentation and size selection steps are not required, decreasing the overall work load and ensuring greater downstream accuracy.

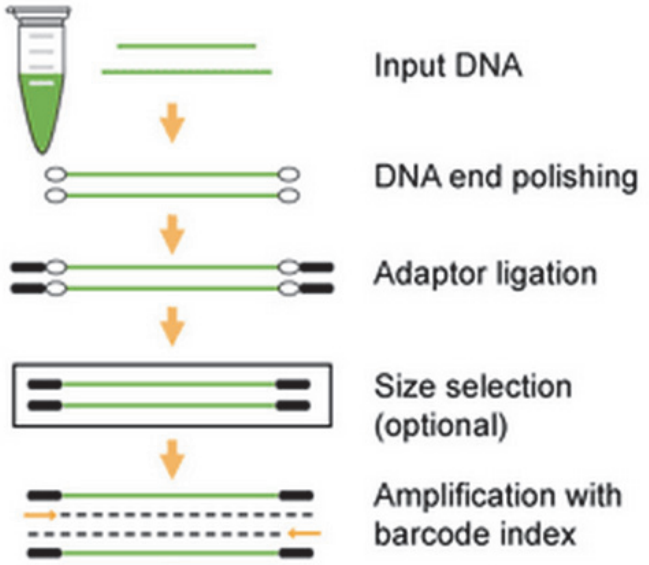


Figure 45: EpiNext Barcode. Digital image. NEB Next Ultra DNA Library Prep Kit. Epigentek, n.d. Web. http://www.epigentek.com/ catalog/ epinext-ngs-barcode-index-set-12-p-3687.html>.

11 PCR reactions were incubated using the same components and thermal recipes previously described, with ratios of unmethylated to methylated dNTPs varying according to the following tables:

Table 8: qPCR Sample Methylation Percentage with Corresponding dNTP Volume.

% Methyl	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
dNTPs (µl)	20	18	16	14	12	10	8	6	4	2	0 µl
5MdNTP(µl)	0	2	4	6	8	10	12	14	16	18	20 µl
Master Mix (µl)	30	30	30	30	30	30	30	30	30	30	30 µl
Total Vol. (µl)	50	50	50	50	50	50	50	50	50	50	50 µl

Table 9: qPCR Master Mix, no dNTPs.

Master Mix Components	**Volume**
GC Reaction Buffer	132 µl
Forward Primer	13.2 µl
Reverse Primer	13.2 µl
Nested Template DNA	13.2 µl
NEB OneTaq Polymerase	3.3 µl
H_2O	221.1 µl
Total Volume	396 µl

The completed reactions were then cleaned up utilizing Ampure beads, and ethanol precipitation. The amplicons were then treated with Lightning Conversion Reagent from the EZ DNA Methylation-Lightning Kit from Zymo Research. This reagent sulphonates non-methylated cytosine nucleotides found in solution, which can then be converted to uracil nucleotides via deamination and desulphonation.

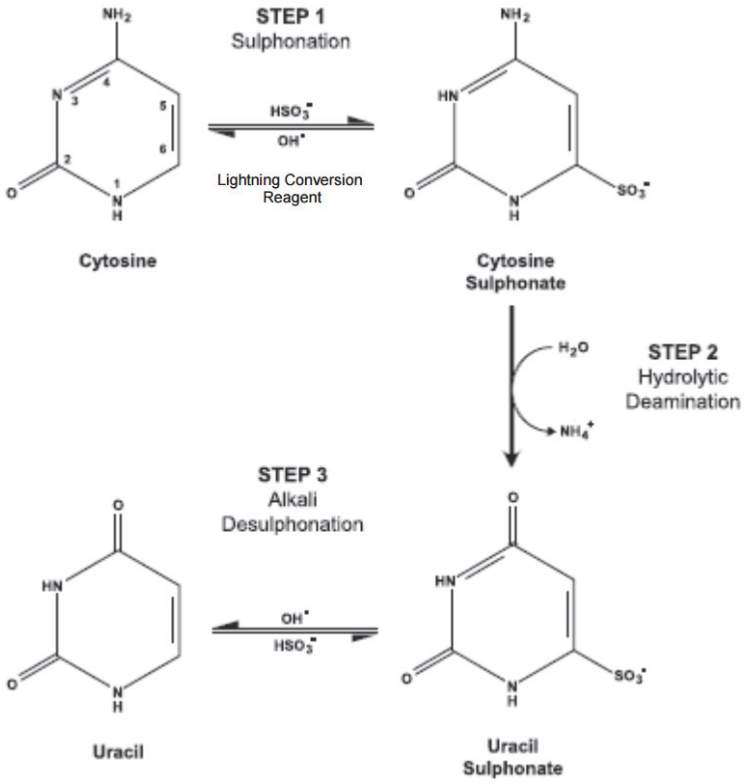


Figure 46: Deamination and Desulphonation Diagram.

Zymo Research Corp. (2013). EZ DNA Methylation-Lightning™ Kit. Irvine, CA: Zymo Research Corp.

20 µl of each cleaned sample was mixed together with 130 µl of the Lightning Conversion Reagent, centrifuged, and incubated at 98 °C for 8 minutes, then 54 °C for 60 minutes. The sulfonated products were then transferred to filtered spin columns, and treated with 600 µl of m-binding buffer and spun at 10,000 g for 30 seconds. The resultant flowthrough was discarded, replaced by 100 µl of m-wash buffer before being spun again for another 30 seconds.

200 µl of desulphonation buffer was then added to each column, which was left to sit for 20 minutes, permitting the reaction to run to completion. Once elapsed, the columns were centrifuged again for 30 seconds, then washed once more with 200 µl of wash buffer, and

subsequent 30 second centrifugation. Finally, 10 µl of elution buffer was added to each column, which were centrifuged a final time for 30 seconds, producing bisulfite treated DNA samples.

To anneal the adaptors for NEBNext, the ends of each amplicon would need to be "polished" so as to present with a 3' adenine tail, and appropriate 5' phosphorylation. This was accomplished by mixing each treated sample with NEBNext End Prep Enzyme Mix, and End Repair Reaction Buffer according to the following table.

Table 10: NEBNext End Prep Recipe.

Components	**Volume**
End Prep Enzyme Mix	3.0 µl
End Repair Reaction Buffer	6.5 µl
Treated DNA (>1 µg)	55.5 µl
Total Volume	65 µl

Each reaction was incubated for 30 minutes at 20 °C, then another 30 minutes at 65 °C. The reactions were then permitted to cool to 4 °C before moving on to adaptor ligation. Ligation was accomplished by adding the following components directly to each reaction.

Table 11: NEBNext Ligation Recipe

Components	**Volume**
Blunt/TA Ligase Master Mix	15 µl
NEBNext Adaptor for Illumina	2.5 µl
Ligation Enhancer	1 µl
Total Volume	83.5 µl

The reactions were mixed, and left to incubate at 20 °C for another 15 minutes, after which 3 µl of USER Enzyme were added to each ligation reaction, in order to cleave the adaptor, leaving a pool of appropriately treated strands. Further treatment was not necessary, as the initial DNA products were all of appropriate length for usage in a DNA library.

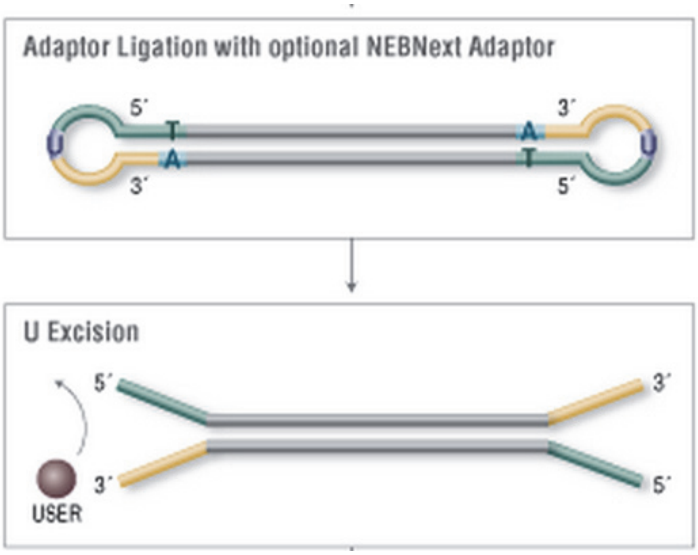


Figure 47: Adapter Ligation. Digital Image. NEB Next Ultra DNA Library Prep Kit. NEB, n.d. Web. https://www.neb.com/products/e7370- nebnext-ultra-dna-library-prep-kit-for-illumina#pd-manuals>.

These titrated methylated library prepped samples could then be run through qPCR. The concentration of each DNA sample was established via uv-vis spectroscopy, and was added together with KAPA SYBR Fast qPCR Master Mix according to the following table.

Table 12: KAPA SYBR Fast qPCR Master Mix Recipe

Component	**Volume**	**Final Concentration**
2X KAPA SYBR Fast qPCR Master Mix	10 µl	1X
10 mM Forward Primer	0.4 µl	0.2 mM
10 mM Reverse Primer	0.4 µl	0.2 mM
Template DNA	Varied	20 ng / 20 µl
H_2O	To 20 µl	N/A
Total Volume	20 µl	N/A

These reactions, along with 6 standard reactions containing a fixed amount of DNA for comparison, were then loaded onto a 96 well plate in triplicate, inserted into a thermocycler, and run according to the following thermal recipe for qPCR.

Table 13: Thermal Recipe for KAPA qPCR

Step	**Temperature**	**Duration**	**Cycles**
Enzyme Activation	95 °C	300 seconds	1 X
Denaturation	95 °C	30 seconds	35 X
Annealing / Extension	60 °C	45 seconds	
Dissociation	95 °C	60 seconds	1 X
	55 °C	30 seconds	
	95 °C	30 seconds	

The resultant amplification data can be seen below.

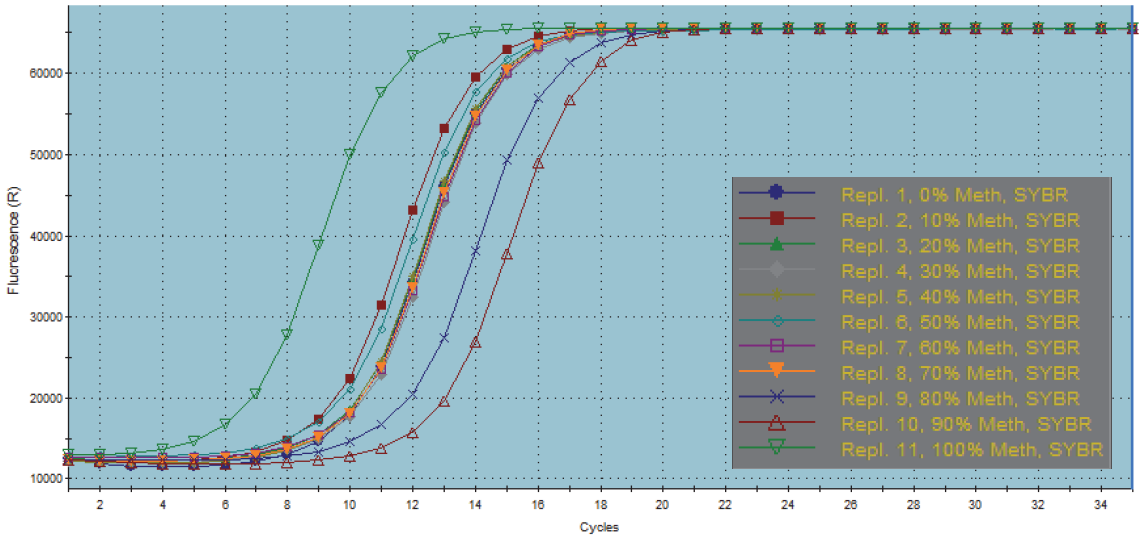


Figure 48: qPCR Methylated Titration Amplification Plot
With the exception of a few outliers at higher methylation percentages, the curvature of each group of samples appears highly uniform, and reaches fluorescent saturation after roughly 12 cycles. This uniformity is to be expected given the utilization of the same amount of starting material, amplifying the same product with varying degrees of methylation.

As can be seen in the figure, the fluorescence curves are all surprisingly uniform, with the majority of samples achieving saturation after roughly 16 cycles. This is not universally the case however. The

reactions containing the greatest concentration of methylated material, 80% 90% and 100% methylated template, all defy this trend, with the former two samples saturating at a slower rate, and the latter doing so much faster. The presence of these outliers can be explained by examining the standard samples. The amplification data for these samples can be seen below.

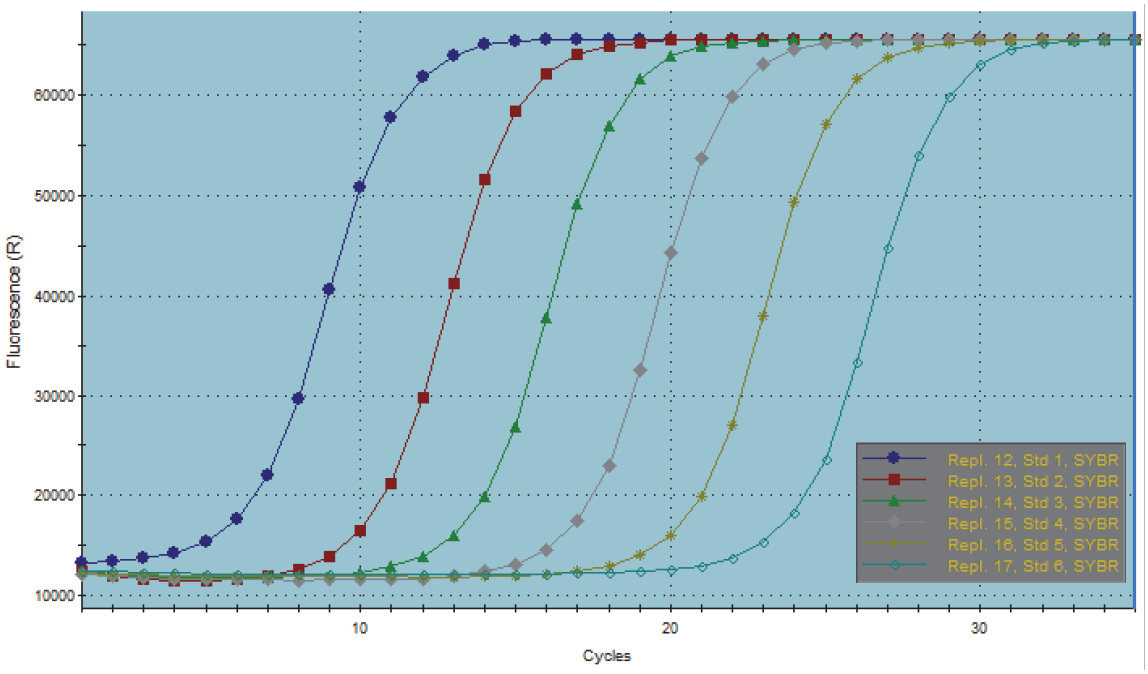


Figure 49: qPCR Methylated Titration Control Standard Amplification Plot. The standards were designed with specific quantities of known material which can be used to create a comparative standard curve. By comparing experimental samples to this curve, the total quantity of generated product can be determined.

Table 14: qPCR Methylated Titration Standards, Quantity Derived from Fluorescence

Replicate	Well Name	Well Type	Quantity
12	Std. 1	Standard	2.00e+001
13	Std. 2	Standard	2.00e+000
14	Std. 3	Standard	2.00e-001
15	Std. 4	Standard	2.00e-002
16	Std. 5	Standard	2.00e-003
17	Std. 6	Standard	2.00e-004

The standards correspond to standardized qPCR reactions with known concentrations and quantities of DNA. This information can be utilized to create a standard curve to help determine the initial quantity of product utilized in each sample. This curve can be seen below.

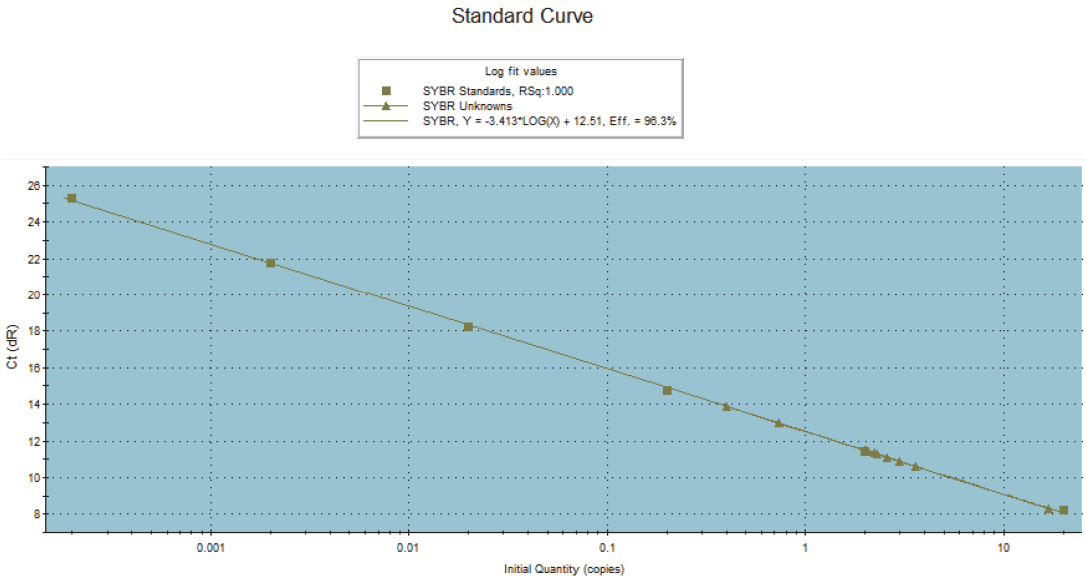


Figure 50 Standard Curve Plot from qPCR Fluorescence Data.
Control standards are shown as squares, and experimental standards are shown as triangles. Samples 9 through 11, with a methylation percentage ranging from 80% to 100%, are the outliers seen between .01 and 1, and beyond 10 copies. The majority are in the range of 2 to 3 copies.

As can be seen in the figure, the majority of samples appear to present with between 2 and 3 initial copies, with two outliers below 1, and a final outlier beyond 10. These specific initial quantities can be seen in the table below.

Table 15: Generated qPCR Methylated Titration Samples, Quantity Derived from Fluorescence

Replicate	**Well Name**	**Well Type**	**Quantity (copies)**
1	0% Meth	Unknown	2.58
2	10% Meth	Unknown	3.60
3	20% Meth	Unknown	2.23
4	30% Meth	Unknown	2.02
5	40% Meth	Unknown	2.32
6	50% Meth	Unknown	2.97
7	60% Meth	Unknown	2.00
8	70% Meth	Unknown	2.07
9	80% Meth	Unknown	0.739
10	90% Meth	Unknown	0.403
11	100% Meth	Unknown	16.87

This indicates a small degree of experimental error in the measurement of the initial concentration of methylated product, and accounts for their deviation from the otherwise uniform threshold cycles. Accounting for this error, it appears that the generation of this library was largely successful. However further analysis of the resultant material will be required for further optimization, and verification.

4.5 – *Conclusion*

While the method here described was successful in the generation of a library of randomly methylated material, there remains much work to be done in order for it to express its full potential. As documented previously, several amplicons containing a high GC percentage (~75%) were notoriously difficult to methylate consistently, and this issue resulted in several downstream experiments involving EMSA and qPCR to focus primarily on alternate control amplicons, in particular B_1.

These difficulties were entirely unexpected, and the process of optimizing experimental procedure in order to overcome this hurdle was further hindered by lack of prior work examining the thermodynamics of integration of methylated dNTPs, and to a much lesser extent, GC rich PCR. Appropriate procedures for generating consistently methylated H/M_1 amplicons were eventually determined, but it is difficult to predict whether such methodology will be applicable to optimizing alternate difficult amplicons. Further work involving a range of epigenetically relevant promoter regions will be required in order to probe this issue, which should be done in tandem with thermodynamic characterization experiments in order to appropriately address this issue. Until this is accomplished, this method can only generate biased randomly methylated DNA libraries, as opposed to libraries containing unbiased, normally distributed methylation.

Additionally, much work will be required to establish the relevance and usefulness of this method in downstream analysis. Integrating this method with further forms of analysis, such as Bisulfite chromatin immunoprecipitation sequencing (BS-ChIP-seq), hydroxyl radical cleavage, and nanopore force spectroscopy, researchers will be able to generate a far more complete picture of any given protein-DNA interaction.

BS-ChIP-seq can be utilized to determine the equilibrium state of protein binding to DNA at various epigenetic states. Hydroxy radical cleavage can reveal conformational structure of methylated DNA *in vivo*, as well as the precise location of protein binding via footprinting. Finally nanopore force spectroscopy enables high-throughput examination of protein-DNA binding strength as compared to a given binding threshold, or on a case-by-case basis via force modulation.

In other words, much further work will be required before this method can be said to have achieved its initial aims. However, given the difficulty of determining the specific mechanisms behind the effects of DNA methylation on gene expression, it appears clear that further work towards reaching this potential is warranted. The rewards of a fully optimized method that can probe the mechanisms behind the impact of methylation on protein-DNA interactions are enormous, and more than justify optimizing this novel method to its full potential.

Appendix:

Primers

Name	Sequence	Primer Length (bp)	Total Length (bp)
HpaII/MspI_1_F	5' – GGTCCTCTTCCTGCTGTCTG – 3'	20	252
HpaII/MspI_1_R	5' – GCCTCCGAGCCTTATAAGGG – 3'	20	252
BamHI_1_F	5' – CCCACTGGTTTGGAGTCTCC – 3'	20	197
BamHI_1_R	5' – ATCCTGTCCACTTCAGCTGC – 3'	20	197
4K_Full_Forward	5' – AACAAACCTGCACATCCTCT – 3'	20	4011
4K_Full_Reverse	5' – CTCATTTCCCTCCATCCTGG – 3'	20	4011
2K_HpaII/MspI_R	5' – TAACCTCAGTGGGTCTCAGT – 3'	20	1929
2K_BamHI_F	5' – ACTGAGACCCACTGAGGTTA – 3'	20	2102
1.1K_HpaII/MspI_F	5' – GCAGATCACCTAAGGTCAGG – 3'	20	1111
1.1K_HpaII/MspI_R	5' – CAGACACATGCTCCTACCTC – 3'	20	1111
1.1K_BamH_F	5' – ACTGAGACCCACTGAGGTTA – 3'	20	1153
1.1K_BamHI_R	5' – GACACCAGAGTCCCAGATTC – 3'	20	1153
0.5K_HpaII/MspI_F	5' – TTTCCTTTCCTCTAAGCGGC – 3'	20	443
0.5K_HpaII/MspI_R	5' – GTACTCACTGGTGGCGAAG – 3'	20	443
0.5K_BamHI_F	5' – AGCAACTGGTGGTTTCTGAT – 3'	20	541
0.5K_BamHI_R	5' – GACACCAGAGTCCCAGATTC – 3'	20	541
H/M_60_F	5' – AAACCATCATGGCGCACACACCT – 3'	23	1181
H/M_60_R	5' – AGAAAAGTTTGCTGGAGCCCGGG –3'	23	1181
H/M_61_F	5' – GGGATCGCAGCGGTCTTAGGGAA –3'	23	827
H/M_61_R	5' – CAGGAAGGAGGCCCGGGGTAAAA–3'	23	827

Bibliography

[11] Tripp, Simon, and Martin Grueber. "Economic Impact of the Human Genome Project." Battelle, 1 May 2011. Web. 4 Feb. 2015.

[2] "Human Genome Project Completion: Frequently Asked Questions." *Human Genome Project Completion: Frequently Asked Questions*. N.p., 30 Oct. 2010.

[3] Lindblad-Toh K, et al. (2011) A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478(7370):476–482.

[4] Ponting CP, Hardison RC (2011) What fraction of the human genome is functional? Genome Res 21(11):1769–1776.

[5] Fraser HB (2013) Gene expression drives local adaptation in humans. Genome Res 23(7):1089–1096.

[6] Grossman SR, et al.; 1000 Genomes Project (2013) Identifying recent adaptations in large-scale genomic data. Cell 152(4):703–713.

[7] Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: Emerging mechanisms and disruption in disease. Am J Hum Genet 76(1):8–32.

[8] Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M (2012) Linking disease associations with regulatory information in the human genome. Genome Res 22(9):1748–1759.

[9] Bernstein BE, Stamatoyannopoulus JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR. The NIH roadmap epigenomics mapping consortium. Nature biotechnology. 2010;28(10):1045-8

[10] Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in the epigenetic domains across cancer types. Nat Genet. 011 Jun 26; 4 3(8):768-75. PMCID: PMC3145050.

[11] Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et Al. Epigenome-wide association data implicate NA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotech. 2013 Feb; 31(2): 142-7.

[12] Spivakov M, Akhtar J, Kheradpour P, Beal K, Girardot C, Koscielny G, et al. Analysis of variation at transcription factor binding sites in Drosophila and humans. Genome biology. 2012 Jan; 13(9):R49.

[13] Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat. Protocols. 2007 Aug; 2(8):1849-61.

[14] Hampshire AJ, Rusling DA, Broughton-Head VJ, Fox KR. Footprinting: A method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. Methods. 2007 Jun; 42(2):128-40.

[15] Capbell CT, Kim G. SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. Biomaterials. 2007 May; 28(15):2380-92.

[16] Lyubchenko YL, Shlyakhtenko LS. AFM for analysis of structure and dynamics of DNA and protein-DNA complexes. Methods. 2009 Mar; 47(3):206-13.

[17] Waddington CH. Introduction to Modern Genetics. London: Allen and Unwin 1939.

[18] Robin Holliday (2006) Epigenetics: A Historical Overview, Epigenetics, 1:2, 76-80, DOI: 10.4161/ epi.1.2.2762

[19] Haig D. The (dual) origin of epigenetics. Cold Spring Harbor Symp Quant Biol 2004; LXIX :1-4

[20] Nanney DL. Epigenetic control systems. Proc Nat Acad Sci USA; 1958; 44:712-7

[21] Russo, V.E.A., Martienssen, R.A., and Riggs, A.D. 1996. Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[22] Yao, Shuyuan, Shuibing Chen, Julie Clark, Ergeng Hao, Gillian M. Beattie, Alberto Hayek, and Sheng Ding. "Long-Term Self-Renewal and Directed Differentiation of Human Embryonic Stem Cells in Chemically Defined Conditions." *Proceedings of the National Academy of Sciences* 103, no. 18 (May 2, 2006): 6907–12. doi:10.1073/pnas.0602280103.

[23] Evans, M. J., and M. H. Kaufman. "Establishment in Culture of Pluripotential Cells from Mouse Embryos." *Nature* 292, no. 5819 (July 9, 1981): 154–56.

[24] Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones. "Embryonic Stem Cell Lines Derived from Human Blastocysts." *Science (New York, N.Y.)* 282, no. 5391 (November 6, 1998): 1145–47.

[25] Ohno, S., W. D. Kaplan, and R. Kinosita. "Formation of the Sex Chromatin by a Single X-Chromosome in Liver Cells of Rattus Norvegicus." *Experimental Cell Research* 18 (October 1959): 415–18.

[26] Avery OT, Macleod CM, McCarty M (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type Iii. J Exp Med 79: 137–158.

[27] McCarty M, Avery OT (1946). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: II. Effect of desoxyribonuclease on the biological activity of the transforming substance. J Exp Med 83: 89–96.

[28] Moore, Lisa D., Thuc Le, and Guoping Fan. "DNA Methylation and Its Basic Function." *Neuropsychopharmacology* 38, no. 1 (2013): 23–38.

[29] Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. Nature Rev. Genet. 10, 295–304 (2009).

[30] Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. Nature Rev. Genet. 9, 465–476 (2008).

[31] Chan, Simon W.-L., Ian R. Henderson, and Steven E. Jacobsen. "Gardening the Genome: DNA Methylation in Arabidopsis Thaliana." *Nature Reviews Genetics* 6, no. 5 (May 2005): 351–60. doi:10.1038/nrg1601.

[32] Rambach A, Tiollais P: Bacteriophage lambda having EcoRI endonuclease sites only in the nonessential region of the genome. Proc Natl Acad Sci U S A 1974, 71:3927-3930.

[33] Suzuki MM, Bird A: DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 2008, 9:465-476.

[34] Riggs AD. X inactivation, differentiation and DNA methylation. Cytogenet. Cell Genet. 1975; 14:9-25.

[35] Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975; 187:226-32.

[36] Pugh JE. Holliday R. Do chemical carcinogens act by altering epigenetic controls through DNA repair rather than by mutations? Heredity 1978; 40:329

[37] Holliday R. A new theory of carcinogenesis. Brit J Cancer 1979; 40:512-3

[38] Nathans, D., and H. O. Smith. "Restriction Endonucleases in the Analysis and Restructuring of Dna Molecules." *Annual Review of Biochemistry* 44 (1975): 273–93. doi:10.1146/annurev.bi.44.070175.001421.

[39] Padmanabhan, R., E. Jay, and R. Wu. "Chemical Synthesis of a Primer and Its Use in the Sequence Analysis of the Lysozyme Gene of Bacteriophage T4." *Proceedings of the National Academy of Sciences* 71, no. 6 (June 1, 1974): 2510–14. doi:10.1073/pnas.71.6.2510.

[40] Primrose SB, Old RW (1994). *Principles of gene manipulation: an introduction to genetic engineering*. Oxford: Blackwell Scientific.

[41] Doerfler W. DNA methylation and gene activity. Ann Rev Biochem 1983; 52:93-124.

[42] Feinberg, A. P., and B. Vogelstein. "Hypomethylation Distinguishes Genes of Some Human Cancers from Their Normal Counterparts." *Nature* 301, no. 5895 (January 6, 1983): 89–92.

[43] Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B: Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet* 1989, 83:155-158.

[44] Haig D. The (dual) origin of epigenetics. Cold Spring Harbor Symp Quant Biol 2004; LXIX: 1-4.

[45] Holliday R. The inheritance of epigenetic defects. Science 1987; 238:163-70

[46] Feinberg, Andrew. "DNA Methylation in Cancer: Three Decades of Discovery." *Genome Medicine* 6, no. 5 (May 30, 2014): 36. doi: 10.1186/gm553.

[47] Cosgrove MS, Wolberger C. How does the histone code work? Biochem. Cell Biol. 2005; 83:468-76

[48] Holliday R. DNA methylation in eukaryotes: 20 years on. In Russo VEA, Riggs AD, Martenssen R, eds. Epigenetic mechanisms of gene regulation. New York. Cold Spring Harbor Laboratory Press ,1996:5-27

[49] Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CI. A genomic sequencing protocol which yield a positivy display of 5-methyl cytosine residues in individual strands. Proc Nat Acad Sci USA 1002; 89:1827-31.

[50] Mayer-Jung, Claudine, Dino Moras, and Youri Timsit. "Effect of Cytosine Methylation on DNA-DNA Recognition at CpG Steps." *Journal of Molecular Biology* 270, no. 3 (July 1997): 328–35. doi:10.1006/jmbi.1997.1137.

[51] Jeltsch A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem 2002; 3:274–93.

[52] Dantas Machado, Ana Carolina, Tianyin Zhou, Satyanarayan Rao, Pragya Goel, Chaitanya Rastogi, Allan Lazarovici, Harmen J. Bussemaker, and Remo Rohs. "Evolving Insights on How Cytosine Methylation Affects protein–DNA Binding." *Briefings in Functional Genomics* 14, no. 1 (January 2015): 61–73. doi:10.1093/bfgp/elu040.

[53] Nathan, Dafna, and Donald M. Crothers. "Bending and Flexibility of Methylated and Unmethylated EcoRI DNA1." *Journal of Molecular Biology* 316, no. 1 (February 8, 2002): 7–17. doi:10.1006/jmbi.2001.5247.

[54] Mirsaidov, U., W. Timp, X. Zou, V. Dimitrov, K. Schulten, A. P. Feinberg, and G. Timp. "Nanoelectromechanics of Methylated DNA in a Synthetic Nanopore." *Biophysical Journal* 96, no. 4 (February 18, 2009): L32–34. doi:10.1016/j.bpj.2008.12.3760.

[55] Fournier, Alexandra, Nobuhiro Sasai, Mitsuyoshi Nakao, and Pierre-Antoine Defossez. "The Role of Methyl-Binding Proteins in Chromatin Organization and Epigenome Maintenance." *Briefings in Functional Genomics* 11, no. 3 (May 2012): 251–64. doi:10.1093/bfgp/elr040.

[56] Defossez, Pierre-Antoine, and Irina Stancheva. "Biological Functions of Methyl-CpG-Binding Proteins." In *Progress in Molecular Biology and Translational Science*, 101:377–98. Elsevier, 2011. http://linkinghub.elsevier.com/retrieve/pii/B9780123876850000123.

[57] Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M. and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39: 457-466.

[58] Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A. and ISSA, J.P. (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet 3: 2023-2036.

[59] G. Egger, G. Liang, A. Aparicio, P.A. Jones Epigenetics in human disease and prospects for epigenetic therapy Nature, 429 (2004), pp. 457–463

[60] W.Y. Chen, S.B. Baylin Inactivation of tumor suppressor genes: choice between genetic and epigenetic routes Cell Cycle, 4 (2005), pp. 10–12

[61] A.P. Feinberg The epigenetics of cancer etiology Semin Cancer Biol, 14 (2004), pp. 427–432

[62] Defossez, P. (2011), "Proteins that Bind Methylated DNA", in Issa, J. (ed.), *DNA Methylation: Physiology, pathology and disease*, The Biomedical & Life Sciences

Collection, Henry Stewart Talks Ltd, London

[63] S.K. Ooi, A.H. O'Donnell, T.H. Bestor Mammalian cytosine methylation at a glance J Cell Sci, 122 (2009), pp. 2787–2791

[64] Dantas Machado AC, Zhou T, Rao S, et al. Evolving insights on how cytosine methylation affects protein–DNA binding. *Briefings in Functional Genomics*. 2015;14(1):61-73. doi:10.1093/bfgp/elu040.

[65] Lazarovici, Allan, et al. "Probing DNA shape and methylation state on a genomic scale with DNase I." *Proceedings of the National Academy of Sciences* 110.16 (2013): 6376-6381.

[66] Collings, Clayton K., Peter J. Waddell, and John N. Anderson. "Effects of DNA methylation on nucleosome stability." *Nucleic acids research* 41.5 (2013): 2918-2931.

[67] Bell, A.C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405: 482-485.

[68] Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M. and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer blocking activity at the H19/Igf2 locus. Nature 405: 486-489.

[69] Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev 15: 490-495.

[70] Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell. 1989; 58:499–507.

[71] Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell.1992; 69:905–14.

[72] Nan, X., Meehan, R.R. and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res 21: 4886-4892

[73] Cross, S.H., Meehan, R.R., Nan, X. and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet 16: 256-259.

[74] Hendrich, B. and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18: 6538-6547.

[75] Fried, M.G. & Crothers, D.M. Equilibria and kinetics of Lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. 9, 6505–6525 (1981).

[76] Hendrickson, W. Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay. BioTechniques 3 (1985) 198–207.

[77] Revzin, A. "Gel electrophoresis assays for DNA-protein interactions." BioTechniques 7.4 (1989): 346-355.

[78] Fried, Michael, and Donald M. Crothers. "Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis." Nucleic acids research 9.23 (1981): 6505-6525.

[79] Fried, Michael G., and Donald M. Crothers. "Kinetics and mechanism in the reaction of gene regulatory proteins with DNA." Journal of molecular biology 172.3 (1984): 263-282.

[80] Barr, H., Hermann, A., Berger, J., Tsai, H.H., Adie, K., Prokhortchouk, A., Hendrich, B. and Bird, A. (2007). Mbd2 contributes to DNA methylationdirected repression of the Xist gene. Mol Cell Biol 27: 3750-3757.

[81] Cross, S.H., Meehan, R.R., Nan, X. and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet 16: 256-259.

[82] Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M. and Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell 105: 487-497.

[83] Woo, H.R., Pontes, O., Pikaard, C.S. and Richards, E.J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev 21: 267-277.

[84] Chen, R.Z., Akbarian, S., Tudor, M. and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet 27: 327-331.

[85] Guy, J., Hendrich, B., Holmes, M., Martin, J.E. and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet 27: 322-326.

[86] Amir, R.E., Van Den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23: 185-188.

[87] Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., Macdougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R. and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science 297: 403-405

[88] Fillion, G.J., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E. and Defossez, P.A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol 26: 169-181.

[89] Sansom, O.J., Berger, J., Bishop, S.M., Hendrich, B., Bird, A. and Clarke, A.R. (2003). Deficiency of Mbd2 suppresses intestinal tumorigenesis. Nat Genet 34: 145-147

[90] Hildebrand, J.D. and Soriano, P. (2002). Overlapping and unique roles for Cterminal binding protein 1 (CtBP1) and CtBP2 during mouse development. Mol Cell Biol 22: 5296-5307

[91] Pingoud, Alfred, and Albert Jeltsch. “Structure and Function of Type II Restriction Endonucleases.” *Nucleic Acids Research* 29, no. 18 (September 15, 2001): 3705–27. doi:10.1093/nar/29.18.3705.

[92] Chu, Da-Chang, Cheng-Keng Chuang, Jin-Bao Fu, Hsien-Siang Huang, Ching-Ping Tseng, and Chien-Feng Sun. "The Use of Real-Time Quantitative Polymerase Chain Reaction to Detect Hypermethylation of the CpG Islands in the Promoter Region Flanking the GSTP1 Gene to Diagnose Prostate Carcinoma." *The Journal of Urology* 167, no. 4 (April 2002): 1854–58.

[93] Chiam, Karen, Margaret M. Centenera, Lisa M. Butler, Wayne D. Tilley, and Tina Bianco-Miotto. "GSTP1 DNA Methylation and Expression Status Is Indicative of 5-Aza-2′-Deoxycytidine Efficacy in Human Prostate Cancer Cells." *PLoS ONE* 6, no. 9 (September 28, 2011): e25634. doi:10.1371/journal.pone.0025634.

[94] Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." *Bioinformatics* 28, no. 12 (June 15, 2012): 1647–49. doi:10.1093/bioinformatics/bts199.

[95] DeAngelis, M M, D G Wang, and T L Hawkins. "Solid-Phase Reversible Immobilization for the Isolation of PCR Products." *Nucleic Acids Research* 23, no. 22 (November 25, 1995): 4742–43.

[96] Sanger, F., and A. R. Coulson. "A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase." *Journal of Molecular Biology* 94, no. 3 (May 25, 1975): 441–48.

[97] De Magalhães, João Pedro, Caleb E. Finch, and Georges Janssens. "Next-Generation Sequencing in Aging Research: Emerging Applications, Problems, Pitfalls and Possible Solutions." *Ageing Research Reviews* 9, no. 3 (July 2010): 315–23. doi:10.1016/j.arr.2009.10.006.

[98] Sanger, F., S. Nicklen, and A. R. Coulson. "DNA Sequencing with Chain-Terminating Inhibitors." Proceedings of the National Academy of Sciences of the United States of America 74, no. 12 (December 1977): 5463–67.

[99] Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent, and L. E. Hood. "Fluorescence Detection in Automated DNA Sequence Analysis." *Nature* 321, no. 6071 (June 12, 1986): 674–79. doi: 10.1038/321674a0.

[100] Cohen, S. N., A. C. Chang, H. W. Boyer, and R. B. Helling. "Construction of Biologically Functional Bacterial Plasmids in Vitro." *Proceedings of the National Academy of Sciences of the United States of America* 70, no. 11 (November 1973): 3240–44.

[101] Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Section 7.1, DNA Cloning with Plasmid Vectors.

[102] Simon, R., U. Priefer, and A. Pühler. "A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria." *Nature Biotechnology* 1, no. 9 (November 1983): 784–91. doi: 10.1038/nbt1183-784.

[103] Klebe, Robert J., et al. "A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast." *Gene* 25.2 (1983): 333-341.

[104] Hanahan, D. "Studies on Transformation of Escherichia Coli with Plasmids." *Journal of Molecular Biology* 166, no. 4 (June 5, 1983): 557–80.

[105] Sidorova, N.Y. and Rau, D.C. 2001Linkage of EcoRI dissociation from its specific DNA recognition site to water activity, salt concentration, and pH: separating their roles in specific and non-specific binding J. Mol. Biol. 310801–816

[106] Sidorova, Nina Y., Shakir Muradymov, and Donald C. Rau. "Trapping DNA–protein Binding Reactions with Neutral Osmolytes for the Analysis by Gel Mobility Shift and Self-Cleavage Assays." *Nucleic Acids Research* 33, no. 16 (January 1, 2005): 5145–55. doi:10.1093/nar/gki808.

[107] Sidorova, N.Y. and Rau, D.C. 2004Differences between EcoRI nonspecific and "star" sequence complexes revealed by osmotic stress Biophys. J. 872564–2576

[108] Engler, L.E., Sapienza, P., Dorner, L.F., Kucera, R., Schildkraut, I., Jen-Jacobson, L. 2001The energetics of the interaction of BamHI endonuclease with its recognition site GGATCC J. Mol. Biol. 307619–636

[109] Kubista M., Andrade JM., Bengtsson M., Forootan A., Jonak J., Lind K., Sindelka R., Sjoback R., Sjogreen B., Strombom L., Stahlberg A., Zoric N., "The real-time polymerase chain reaction", Mol Aspects Med. 27(2-3):95-125, 2006.

[110] Holland P., Abramson R.D., Watson R. and Gelland D.H., "Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of thermos aquatus.", Proc. Natl, Acad, Sci. USA, Vol.88, 7276-7280, 1991

[111] Carr, A. C.; Moore, S. D. (2012). Lucia, Alejandro, ed."Robust Quantification of Polymerase Chain Reactions Using Global Fitting". *PLoS ONE* 7 (5): e37640.
[112] Pingoud, Alfred, and Albert Jeltsch. "Structure and Function of Type II Restriction Endonucleases." *Nucleic Acids Research* 29, no. 18 (September 15, 2001): 3705–27. doi:10.1093/nar/29.18.3705.

[113] Mao, Fei, Wai-Yee Leung, and Xing Xin. "Characterization of EvaGreen and the Implication of Its Physicochemical Properties for qPCR Applications." *BMC Biotechnology* 7, no. 1 (November 9, 2007): 76. doi:10.1186/1472-6750-7-76.

[114] Dwight, Zachary, Robert Palais, and Carl T. Wittwer. "uMELT : Prediction of High-Resolution Melting Curves and Dynamic Melting Profiles of PCR Products in a Rich Web Application." *Bioinformatics*, February 7, 2011, btr065. doi:10.1093/bioinformatics/btr065.

[115] Lefebvre,A., Mauffret,O., Antri,S.E., Monnot,M., Lescot,E. and Fermandjian,S. (1995) Sequence dependent effects of CpG cytosine methylation. A joint 1H-NMR and 31P-NMR study. Eur. J. Biochem., 229, 445–454

Curriculum Vitae:

Gabriel Deards was born in New York in 1990. He received bachelor's degrees in Biophysics, and Statistics concentrated in Quantitative Biology from The Macaulay Honors College at Hunter in 2013. During his undergraduate career, he conducted research with the Goddard Lab at Hunter College, the Austin Group at Princeton University, and CLUE at Columbia University. Following graduation, Gabriel matriculated at Johns Hopkins University, pursuing a Master's degree in Biomedical Engineering. Gabriel conducted his research while working in the Timp Lab, under Professor Winston Timp, as part of Johns Hopkins University's Master's degree requirements in Biomedical Engineering.